DOCUMENT RESUME

ED 052 633 24 EM 009 113

AUTHOR Johnson, Mauritz; Posner, George J.
TITLE Testing the Effect of Verbal-Quantitative Aptitude Discrepancy On the Learning of Deductive Reasoning Through Programmed Instruction. Final Report.
INSTITUTION State Univ. of New York, Albany. Research Foundation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-9-B-054
PUB DATE Feb 71
GRANT OEG-2-9-520054-1033(010)
NOTE 65p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Ability Grouping, Abstract Reasoning, Aptitude, *Critical Thinking, Individual Differences, Individualized Instruction, *Intermode Differences, Logical Thinking, Low Ability Students, *Nonverbal Ability, *Programed Instruction, Symbolic Learning, *Verbal Ability

ABSTRACT

An experiment was conducted to investigate three topics of current interest in education: critical thinking, programed instruction, and aptitude-treatment interaction. Thirty college students and 31 ninth graders with extreme discrepancies between their verbal and quantitative aptitude scores were taught eight principles of class reasoning by random assignment to one of two versions of a programed instruction unit. One version contained verbal examples and scientific terminology; the other substituted letter symbols for the scientific words. It was hypothesized that if adaptation of instruction to individual differences is an efficacious practice, then students who were highly verbal would learn better from the verbal program version and students with low verbal aptitude would benefit from the symbolic version. Although there was a slight tendency in the hypothesized direction, in no instance was the interaction between program version and discrepancy type significant at the .05 level. However, it was demonstrated that class reasoning principles can be taught by means of a relatively short instructional program and that such instruction will transfer to understanding of conditional reasoning principles. (Author/JY)

FINAL REPORT
Project No. 9-B-054
Grant No. OEG-2-9-420054-1033 (010)

TESTING THE EFFECT OF VERBAL-QUANTITATIVE APTITUDE DISCREPANCY ON THE LEARNING OF DEDUCTIVE REASONING THROUGH PROGRAMMED INSTRUCTION

Mauritz Johnson, Principal Investigator
and
George J. Posner
State University of New York at Albany
The Research Foundation of State University of New York
P.O. Box 7126
Albany, New York 12224

February, 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report

Project No. 9-B-054
Grant No. OEG-2-9-420054-1033 (010)

TESTING THE EFFECT OF VERBAL-QUANTITATIVE APTITUDE DISCREPANCY ON THE LEARNING OF DEDUCTIVE REASONING THROUGH PROGRAMMED INSTRUCTION

Mauritz Johnson, Principal Investigator,
and
George J. Posner

State University of New York at Albany

Albany, New York

February, 1971

The research reported herein was performed pursuant to a grant 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.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Table of Contents

List of Tables

I

Summary of the Study

Samples consisting of thirty college students and 31 ninth graders with extreme discrepancy between their verbal and quantitative aptitude scores were taught eight principles of class reasoning by random assignment to one of two versions of a programmed instruction unit, one version containing verbal examples using scientific terminology, the other substituting letter symbols for the scientific words. The Cornell Class Reasoning and Conditional Reasoning Tests were administered before and after instruction to measure direct learning of class reasoning principles and transfer to principles of conditional reasoning. Interaction between direction of aptitude discrepancy and program version was the chief object of interest. It was hypothesized that if adaptation of instruction to individual differences is an efficacious practice, then both for direct learning and transfer, students with high verbal - low quantitative aptitudes should perform better after instruction with the verbal program version, students with high quantitative - low verbal aptitudes should benefit most from the symbolic version, and students of both discrepancy types should gain more from a program matching their type than from a mis-match.

The results of the experiments revealed that in 20 out of 26 comparisons made there were slightly superior gains in the hypothesized direction, but in no instance was the interaction between program version and discrepancy type significant at the .05 level. It was surmised that the instructional program was not long enough, the two versions were not different enough, and in view of the large within-group variance, the groups were not large enough to produce a significant interaction.

Nevertheless, it was demonstrated that class reasoning principles can be taught by means of a relatively short instructional program and that such instruction will transfer to understanding of conditional reasoning principles. In addition, tentative norms for college students based on a sample of 108, were established for both of the Cornell Reasoning tests, demonstrating that they can be used at the college level and, although some inconsistency was evident in the data, that little if any growth in reasoning ability occurs during the college years, whether or not students take courses in logic. No significant sex differences in reasoning ability were found.

Both as to percent demonstrating mastery of each principle and as to the relative difficulty of the principles, the results of this study were generally consistent with those reported by Robert H. Ennis and Dieter Paulus in Cooperative Research Project No. 1680, Critical Thinking Readiness in Grades 1-12 (Phase 1: Deductive Logic in Adolescence), 1965.

II

Context of the Study

The study reported herein bears upon three topics of great current interest in education: critical thinking, programed instruction, and aptitude-treatment interaction. In addition it deals with a fourth concept that has received virtually no attention, namely, aptitude discrepancy. This discrepancy between an individual's verbal and quantitative aptitude served in this study as the personological variable interacting with programed instruction formats aimed at teaching critical thinking.

Critical Thinking

That students should learn to think in school is hardly a new proposition, but its importance has recently been emphasized in such reports as those on the central purpose of American education (Educational Policies Commission, 1961) and on logical operations in the classroom (Smith and Meaux, n.d.). Robert H. Ennis' definition of critical thinking as "the correct assessment of statements" (Ennis, 1961) has served as a basis for several studies of children's readiness to learn principles of deductive reasoning (Ennis and Paulus, 1965; Ennis et al., 1969) and for a recent book, Logic in Teaching (Ennis, 1969). Among the "process objectives" currently receiving considerable professional attention (Parker and Rubin, 1966; Rubin, 1969; Berman, 1968; Andreas, 1970), thinking skills constitute a major type of "transcurricular" objective, pertinent to most, if not all, subject matter fields. While there are other thinking skills besides those of deductive reasoning or logic and more to deduction than the application of principles of conditional and class reasoning, these two forms of logic are of great significance and wide applicability. The present study provides data supplementing some of Ennis' findings regarding the reasoning ability of secondary-school students and, more significantly, extending them to the college level.

Programed Instruction

The use of programed instruction in the teaching of logic is not novel. A number of programs are available commercially (e.g., Dickoff and James, 1965; Schagrin, 1968). Logic lends itself well to programing, and the lack of confidence on the part of most teachers concerning their ability to teach deductive reasoning suggests that the use of the programed mode for this purpose would be particularly appropriate. Sharlow (1971) has offered evidence, however, that students are more likely to apply logic learned through programed instruction to their work in subject areas, such as mathematics, when specifically encouraged and helped to do so by the teacher of the subject.

But whether or not logic can be taught through programed instruction is not an issue, for clearly it can be. What is of greatest research interest is what program characteristics are most associated with effective learning and transfer by students with various characteristics. Early interest in such program characteristics as linear vs. intrinsic format, error rate, step size, nature of prompts, types of reinforcement, and response modes has given way to concern with such features as semantic, symbolic, and figural presentation modes. In the present study program versions differed with respect to the extent of symbolic and verbal semantic material employed.

Aptitude-Treatment Interaction

The period of time during which this study was conducted saw a marked increase in interest in aptitude-treatment interaction (ATI) and the appearance of several highly significant articles on the subject. Although the entire notion of individualizing instruction rests upon the assumption that differential instructional treatment with respect to learner characteristics is both desirable and effective, very few studies have demonstrated that any such interaction exists. Cronbach (1967) stressed the importance of investigating interactions between various characteristics of learners and instructional treatments. Writing with Snow, he made the assertion that, "It is inconceivable to us that humans, differing in as many ways as they do, do not differ with respect to the educational treatment that fits each one best." Nevertheless, Cronbach and Snow consider ATI studies to be "high-risk research" (Cronbach and Snow, 1969, p. 193). Bracht (1969) has suggested the advisability of employing extreme variations in learner characteristics in seeking evidence in interaction. Bracht (1969; 1970) analyzed 108 ATI studies and found only five that demonstrated interaction, a percentage which might be predicted to occur due to chance under the customary .05 fiducial limit. Bracht added, on the basis of empirical evidence from analyzed studies, three further necessary, but not sufficient, conditions for the demonstration of aptitude-treatment interaction. These conditions are: (1) a controlled, rather than uncontrolled, treatment, (2) a factorially simple, rather than complex, personological variable, and (3) a specific, rather than general, dependent variable (Bracht, 1970). The present study entailed a controlled treatment and a specific dependent variable. Whether aptitude discrepancy is factorially simple or complex is a matter for conjecture. Little attention has been paid to it.

V-Q Aptitude Discrepancy

Bracht (1970, p. 638) makes the observation that, while such personological variables as "general ability" and "previous achievement" are factorially complex, an under-over achievement variable derived from them may be factorially simple. By analogy, one might make the conjecture that, while "general scholastic aptitude", derived from a combination of verbal and quantitative aptitude scores, is indeed factorially complex, the discrepancy between verbal and quantitative aptitude may be factorially simple, extreme discrepancies in one direction or the other defining clearly distinct personological types.

The well-recognized "positive correlation of traits" leads to the prediction that "V-Q aptitude discrepancy" will in most instances be relatively small, with central tendency of zero. Individuals with expremely high or low scholastic aptitude necessarily have small V-Q aptitude discrepancy. Individuals with average scholastic aptitude, on the other hand, differ greatly with respect to V-Q aptitude discrepancy. The most extreme instances of discrepancy fall at the mean of the total aptitude continuum. Using T-scores, the extremely discrepant 80-20 and 20-80 cases have the same total aptitude as the non-discrepant 50-50 cases. Yet these ostensibly "average" individuals clearly represent quite divergent characteristics. Presumably, a student with a V-score of 80 and Q-score of 20 responds quite differently to a given set of instructional materials than one with a V-score of 20 and a Q-score of 80. Whether he does in fact was what this study primarily sought to investigate. In a companion study (Johnson and Posner, in progress) the investigators studied the characteristics of individuals with marked aptitude discrepancies and the concomitants and etiology of aptitude discrepancy.

III

Purposes and Hypotheses

The primary objective of this study was to determine whether learning is facilitated by the use of instructional materials that are consistent with significant characteristics of the learner. The characteristic considered in this study was differential scholastic aptitude. Subjects for the study were students with maximum discrepancies between their verbal and quantitative aptitude scores. By definition, students with high scores in both verbal and quantitative aptitude may be expected to learn better than those whose scores in both types of aptitude are low. But it is not evident that students with extremely high verbal, and correspondingly low quantitative, aptitudes will in general perform better or worse in a learning situation than students with extremely high quantitative, and correspondingly low verbal, aptitudes. The first of these two groups may learn more readily than the other group when one kind of instructional material is used and less well with a different kind. It was this possibility that was examined in the present study.

Most instructional materials are verbal in nature. Such materials would appear to be quite suitable for students with high verbal aptitudes, but may present difficulty for those whose verbal aptitudes are low. If the latter have extremely high quantitative aptitudes, it might be reasonable to assume that their learning would be facilitated by materials which made greater use of symbols and minimized verbal content. To test this assumption it would be necessary to find or develop materials which deal with the same content and are as similar as possible except with respect to their verbal or symbolic format. For the present study, two linear instructional programs were developed, both designed to teach the same principles of deductive reasoning, but differing in that one version was entirely composed of words, whereas the other employed symbols as much as possible. Obviously, it is impossible to eliminate words completely. The only difference between the two versions was that one provided little contact with symbols and the other emphasized them.

The eight principles of class reasoning which were taught by the program were the ones used by Ennis and Paulus, in developing The Cornell Class Reasoning Test (Ennis and Paulus, 1965, pp. II-12-13). They are as follows:

1. Whatever is a member of a class is not a non-member of that class and vice versa.
2. Whatever is a member of a class is also a member of a class in which that class is included.

3. Whatever is a member of a class is not necessarily a member of a class included in that class.
4. Class exclusion is symmetric.
5. Whatever is a member of a class is not a member of a class excluded from the first.
6. Whatever is not a member of a class is not necessarily also not a member of a class in which the first class is included.
7. Whatever is not a member of a class is not necessarily a member of (nor a non-member of) another class which is excluded from the first class.
8. Whatever is not a member of a class is also not a member of any class included in the first class.

Given, then, two groups of students with opposite types of extreme aptitude discrepancy and a program version presumed favorable to each group, the hypothesis to be tested was that widely assumed in the pedagogical literature, namely, that each group would learn more effectively when matched with the appropriate type of program. A second hypothesis was that matching student attributes with instructional treatment would facilitate not only direct learning but transfer of learning as well.

Experimental Hypotheses

The two aptitude discrepancy groups were designated Vq and Qv to represent, respectively, students with high verbal and low quantitative aptitude scores and students with high quantitative and low verbal scores. The two program versions were designated SY and VS, the first emphasizing the use of symbols, the second being purely verbal. The "S" in VS indicated that the illustrative subject matter to which the logical principles were applied in this version was scientific. Other verbal programs could be developed using subject matter from other fields, such as the humanities. The SY version, of course, had no illustrative subject matter at all, since all technical words had been replaced by literal symbols.

The dependent variables were scores on two reasoning tests developed at Cornell University. These tests will be discussed later, but one measured performance in class (categorical) reasoning and the other, performance in conditional (hypothetical) reasoning. Since the instructional programs attempted to teach class reasoning, the first test (designated CL) provided a measure of direct learning as s dependent variable. The other test, dealing with conditional reasoning and designated CO, was the measure of transfer of learning as a dependent variable.

The experimental hypotheses can now be stated in theoretical, operational, and null forms, as follows:

A. General theory: Adaptation of instructional materials to learner characteristics facilitates learning.

Theoretical hypothesis 1: Students with high verbal aptitude and low quantitative aptitude will learn principles of class reasoning better from verbal instructional materials than from materials emphasizing symbols.

Operational hypothesis (H_1): Students in group Vq who complete program VS will score higher on test CL than students in group Vq who complete program SY.

Null hypothesis (H_0): The probability exceeds .05 that a difference in scores on test CL as large as that by which the scores of Vq students completing program VS exceed the scores of Vq students completing program SY would occur by chance.

Theoretical hypothesis 2: Students with high quantitative and low verbal aptitude will learn principles of class reasoning better from instructional materials emphasizing symbols than from verbal materials.

Operational hypothesis (H_2): Students in group Qv who complete program SY will score higher on test CL than students in group Qv who complete program VS.

Null hypothesis (H_0): The probability exceeds .05 that a difference in scores on test CL as large as that by which the scores of Qv students completing program SY exceed the scores of Qv students completing program VS would occur by chance.

Theoretical hypothesis 3: Students with marked discrepancies between their verbal and quantitative aptitudes will learn principles of class reasoning better from instructional materials consistent with their higher aptitude than from materials consistent with their lower aptitude.

Operational hypothesis (H_3): Students in group Vq who complete program VS and students in group Qv who complete program SY will score higher on test CL than students in group Vq who complete program SY and students in group Qv who complete program VS.

Null hypothesis (H_0): Designating those Vq students who complete program VS and those Qv students who complete program SY as group M (matched) and designating those Vq students who complete program SY and those Qv students who complete program VS as group U (mismatched), the probability exceeds .05 that a difference in scores on test CL as large as that by which the scores of group M exceed the scores of group U would occur by chance.

The above operational and null hypothesis may be more conveniently stated in symbolic form, as follows:

H_1: $(Vq \cdot VS)_{CL} > (Vq \cdot SY)_{CL}$
H_0: $p\ [(Vq \cdot VS)_{CL} - (Vq \cdot SY)_{CL}] > .05$

H_2: $(Qv \cdot SY)_{CL} > (Qv \cdot VS)_{CL}$
H_0: $p\ [(Qv \cdot SY)_{CL} - (Qv \cdot VS)_{CL}] > .05$

$M = (Vq \cdot VS) + (Qv \cdot SY)$
$U = (Vq \cdot SY) + (Qv \cdot VS)$

H_3: $(M)_{CL} > (U)_{CL}$
H_0: $p\ [(M)_{CL} - (U)_{CL}] > .05$

B. General theory: Adaptation of instructional materials to learner characteristics facilitates transfer of learning.

Theoretical hypothesis 4: Students with high verbal aptitude and low quantitative aptitude will transfer principles of class reasoning to problems in conditional reasoning better when taught class reasoning through verbal instructional materials than through materials emphasizing symbols.

This and subsequent theoretical hypotheses can be most conveniently operationalized through use of the symbols employed earlier. These are:

- Vq - students with high verbal and low quantitative aptitude scores
- Qv - students with high quantitative and low verbal aptitude scores
- VS - a linear program in a verbal format designed to teach certain principles of class reasoning
- SY - a similar linear program emphasizing the use of symbols
- Vq·VS - students in group Vq taught through program VS
- Vq·SY - students in group Vq taught through program SY
- Qv·VS - students in group Qv taught through program VS
- Qv·SY - students in group Qv taught through program SY
- M - students in groups Vq and Qv taught through program most consistent with their higher aptitude (Vq·VS and Qv·SY).
- U - students in groups Vq and Qv taught through the program least consistent with their higher aptitude (Vq·SY and Qv·VS).
- CL - subscript indicating scores on class reasoning test
- CO - subscript indicating scores on conditional reasoning test.

Operational hypothesis (H_4): $(Vq \cdot VS)_{CO} > (Vq \cdot SY)_{CO}$
Null hypothesis (H_0): $p\ [(Vq \cdot VS)_{CO} - (Vq \cdot SY)_{CO}] > .05$

Theoretical hypothesis 5: Students with high quantitative and low verbal aptitude will transfer principles of class reasoning to problems in conditional reasoning better when taught class reasoning through instructional materials emphasizing symbols than through verbal materials.

H_5: $(Qv \cdot SY)_{CO} > (Qv \cdot VS)_{CO}$
H_0: $p\,[(Qv \cdot SY)_{CO} - (Qv \cdot VS)_{CO}] > .05$

Theoretical hypothesis 6: Students with marked discrepancies between their verbal and quantitative aptitudes will transfer principles of class reasoning to problems in conditional reasoning better when taught class reasoning through instructional materials consistent with their higher aptitude than through materials consistent with their lower aptitude.

H_6: $(M)_{CO} > (U)_{CO}$
H_0: $p\,[(M)_{CO} - (U)_{CO}] > .05$

Secondary Objective

Since this study entailed the teaching of class reasoning and the testing of college students in both class and conditional reasoning, it was a subsidiary objective to obtain information regarding college students' ability to reason and whether growth in this ability occurs during the college years. The developer of the Cornell Reasoning Tests has reported (Ennis and Paulus, 1965; Ennis, 1969) tentative norms for each grade in the lower schools, but not for the college years. Although the testing done in the present study was confined to one institution of higher learning, data were obtained that shed some light on the question of whether most students have already mastered the principles of deductive reasoning tested by the Cornell tests by the time they enter college, and if not, whether their reasoning ability improves while in college.

Limitations and Assumptions of the Study

As in every study, a number of assumptions were made in this one which, together with certain features of the design, may limit the generalizability of the results. The most significant of these assumptions and features are identified here to permit the reader to judge what cautions need to be exerciesd in generalizing the findings.

1. Population. The subjects who participated in the study were restricted to selected students at one medium-sized eastern state university and selected ninth graders in one school district in New York State. The assumption is that any interaction between instructional treatment and aptitude discrepancy is

independent of geographic region and characteristics of the educational institution attended.

2. Sample size and selection. The principal experimental data were obtained on a total sample of 62 students who either volunteered or agreed, with parental permission, to participate, and in the case of the college sample, were paid for doing so. The assumptions are that the statistical procedures employed in analyzing the data appropriately took into account the size of the sample and that the extent to which the sample was unrepresentative of students with extreme aptitude discrepancies did not affect the aptitude-treatment interaction (ATI).

3. Program length and effectiveness. The two linear program versions each consisted of about 200 frames. It is possible that a longer program of different design might have been more effective in teaching class reasoning. The assumptions are that no minimum amount of learning is essential before an aptitude-treatment interaction occurs and that the programs used would produce a sufficient amount of learning to make tha ATI discernable with the tests employed if such an interaction existed. It was not assumed that the programs had maximum, or even optimum, effectiveness, nor that the two versions were equally effective.

4. Aptitude-program consistency. The only adaptation of the verbal program was the replacement of specific scientific words by letter symbols. The assumption is that quantitative aptitude involves, in part at least, a facility in manipulating symbols other than words. It was also assumed that verbal aptitude includes facility with words related to science, but not that it is limited to such words.

5. Content. The instructional treatment was limited to the teaching of class reasoning. The assumption is that ATI is not specific to the type of cognitive learning outcome. Class reasoning was selected because (1) it is a non-trivial learning outcome, (2) it is an ability that is widely applicable to many fields of learning, (3) it lends itself well to programed instruction, and (4) convenient tests of achievement and transfer were available.

6. Programed instruction. The instructional treatment was limited to the programed mode. The assumption is that ATI is not limited to any particular mode of instruction, such as so-called "conventional

classroom instruction involving teacher-pupil interaction." Programed instruction was selected because (1) it standardizes the treatment by circumventing the teacher variable, (2) it permits flexibility in scheduling the treatment individually, (3) it allows each subject all the instructional time he requires; and (4) it results in permanently available programs for future use.

7. Tests. The validity and reliability of the measures of the dependent variables are dependent on those of the Cornell Reasoning Tests, which are reported in a later section. The assumption was that these tests were suitable measures of what was learned, since the tests and programs dealt with the same principles of reasoning.

8. Testing. The same test forms were used as pre-test and post-test. The assumption was that the nature of the 72 items was such that little memory effect would be present after a period of several weeks, particularly since subjects were not informed as to total score or items correct on the pre-test.

9. Program administration. Subjects were permitted to complete the instructional programs at their own pace without supervision, but were required to respond overtly to each frame and to turn in programs when completed. The assumption is that the programs were in fact completed by the subjects themselves and, further, that any deviation from this assumption was randomly distributed across program format and attribute classification.

10. Norming. In the non-experimental, norming portion of the study, students whose names were selected from a directory were invited to take one or both tests and were offered remuneration for doing so. The assumptions are that the selection procedure, though systematic, was unbiased and that the prospect of remuneration did not affect the nature of the volunteers.

11. Scoring. Test responses for both the experimental and norming portions of the study were recorded by the subjects on separate answer sheets and scored by hand using an answer key overlay. The assumptions are that the answer key was correct, that there were no systematic errors in responding or scoring, and that any non-systematic errors in either process were randomly distributed across the experimental groups.

12. Classification. The college and ninth grade subjects were classified as having aptitude discrepancies on the basis of recorded scores on a single test at each level. The assumptions are that the test appropriately measured the aptitudes in question and were accurately scored and that the results were correctly recorded.

IV

Research Design and Procedures

The major experimental activities carried out in the study were the identification of potential subjects, pre-testing, development and administration of instructional programs, and post-testing. There were slight variations in these procedures between the college and middle school experiments. The college procedures are discussed first.

Identification of Subjects

To determine the feasibility of finding suitable subjects with extreme aptitude discrepancies, an analysis was made of the distribution of scores on the scholastic aptitude portion of the New York State Regents Scholarship Examination (RSE) submitted for college admission by students at the State University of New York at Albany. The verbal section of this test, dealing with "same-opposites", "verbal analogies", and "sentence completion", has a total possible score of 100. The maximum score on the quantitative section is 50. Since the two scores are moderately highly correlated, most students' scores are not very discrepant. Furthermore, the lower a student's higher score is, the smaller the discrepancy can be, regardless of how low the lower score is. The problem in identifying those with the greatest discrepancy is to set as high a minimum higher score, and as low a maximum lower score, as possible without eliminating virtually every potential subject.

Three other factors affect aptitude discrepancy among college students. One is that, in relation to the full range of scores achieved by those taking the scholastic aptitude test, there will be in any college population very few instances of extremely low scores for either aptitude and an abundance of instances of extremely high scores for each aptitude.

But the admissions selection process has a second effect. It creates a situation wherein the probability that a low score is associated with discrepancy is considerably greater than the probability that a high score is. That is to say, for a student with one very low aptitude score, the other score must have been fairly high, or he probably would not have been accepted. But for a student with one very high score, there is no corresponding demand that his other score be below any given point.

Finally, there is the fact that greater attention is usually given, and probably properly so, to the verbal aptitude score than to the quantitative. Relatively speaking, applicants with a low verbal score are not as likely to be admitted to college as applicants with a low quantitative score. It follows that for any given discrepancy, there are likely to be more students of the Vq type than of the Qv type (capital letter denotes higher aptitude).

A list of potential subjects for this study was obtained through a computer printout of all undergraduate students except Freshmen registered as SUNY-Albany whose RSE scores met one of the two following criteria:

Vq: V score > 83 and Q score < 30
Qv: V scores < 62 and Q score > 43

Out of a total initial enrollment of 5731 in the four classes, the application of these criteria identified 220 individuals or 3.9 percent. They were distributed by class and discrepancy type as shown in Table 1.

Table 1

Distribution of Students With Marked Verbal-Quantitative Aptitude Discrepancy at State University of New York at Albany

Class	Discrepancy Type: Vq	Qv	Total	Total in Class	Percent Discrepant
1969	27	18	45	1312	3.4
1970	21	30	51	1240	4.1
1971	38	32	70	1472	4.8
1972	32	22	54	1707	3.2
Total	118	102	220	5731	3.9

By the time the development of instructional programs had been completed, the Class of 1969 had graduated. Letters were sent to 115 "discrepant" students in the remaining three classes for whom college addresses were available explaining in general the purpose of the study and inviting them to appear at a group testing session to take the pre-test in class reasoning, or to take it individually at their convenience. Of the 115 potential subjects, 43 took the test, for which each received a stipend of two dollars. Of the 43 that took the pre-test 11 were eliminated since they had attained a score above 67. All 32 of the remaining subjects consented to continue for an additional stipend of fifteen dollars, although two of these subjects did not complete the study, due to disruptions on campus.

The median verbal and quantitative aptitude scores of the thirty participants are shown in Table 2, broken down by their assignments to program versions. The lowest verbal score of the Vq group exceeds by more than twenty points the highest verbal score of the Qv group. Similarly, the lowest quantitative score of the Qv group is twenty points higher than the highest quantitative score of the Vq group. The Vq-VS group turned out to be somewhat less discrepant than the other three.

Table 2

Verbal and Quantitative Aptitude Scores of College Students Participating in Study

		Verbal Aptitude		Quantitative Aptitude[1]		Discrepancy	
Group	N	Median	Range	Median	Range	Median	Range
Vq-VS	8	83.0	83-90	54.0	28-66	29.5	17-55
Vq-SY	7	85.0	83-88	48.0	28-58	38.0	26-55
Total Vq	15	84.0	83-90	52.0	28-66	34.0	17-55
Qv-VS	7	53.0	42-59	90.0	86-100	40.0	36-44
Qv-SY	8	55.0	50-61	91.0	86-98	35.0	26-42
Total Qv	15	54.0	42-61	90.0	86-100	37.0	26-44

[1] Maximum V-score is 100, maximum Q-score is 50. Q-scores here have been doubled to permit discrepancy calculation.

Pre-testing

In addition to serving as co-variate controls, the scores on the class reasoning pre-test provided a basis for selecting subjects for participation in the study. It was desired to exclude students who scored too high on the pre-test to be able to show any improvement as a result of instruction

The Cornell Class Reasoning Test (Form X) was developed by Robert H. Ennis and his associates in 1964 for use in their studies of readiness to learn critical thinking. The test consists of 72 items, comprising twelve groups of six items each. Nine of the groups test eight principles of class reasoning, one group of six items testing each of seven principles and two groups (twelve items) testing Principle 3. The three remaining item groups test combinations of two or three principles.

Each group is comprised of six items, representing three types, designated CF (concrete familiar), SY (symbolic), and SU (suggestive). The four CF items in a group mention concrete articles and qualities which are familiar to the subject and in statements which he has no reason to believe are true or false. The one SY item substitutes letter symbols for the concrete familiar terms. The remaining item (SU) has familiar content but the empirical truth status of statements is known to the subject and is inconsistent with their validity.

The six items in each of the twelve item groups are scattered throughout one half of the test or the other. Six practice items instruct the subject how to respond. In each item, he is first given one or more premises, prefaced by the phrase, "Suppose you know that..." He is then presented with a possible conclusion, prefaced by, "Then would this be true?" The alternative responses in each instance are "Yes", "No", and "Maybe". The instructions indicate that if, and only if, the proposed conclusion must be true, i.e., necessarily follows from the premises, the proper response is "Yes". Similarly, if and only if, the proposed conclusion must be false, i.e., could not follow from the premises, the response should be "No". If the proposed conclusion does not necessarily follow from the premises, but also is not ruled out be them, then the appropriate response is "Maybe". The test booklets provide for circling the alternative selected; in this study a separate answer sheet was used, on which the subjects placed an X over the letter corresponding to the selected response. Scoring was done by hand with the use of an overlay key.

The maximum possible total score on the test is 72 Ennis, however, is more concerned with how many principles are mastered and which ones. He defines the necessary and sufficient criterion of "mastery of a principle" as a minimum of five correct responses out of six. "Failure to master a principle" is defined as three or fewer correct responses. Four correct responses is indeterminate--neither mastery nor failure to master. (Ennis and Paulus, 1965, p. IV-15).

One might desire that subjects for the experiment not have mastered more than half of the principles. On the 54 items dealing with single principles, it would be possible for a score as low as 25 to represent mastery of more than half of the principles (5 principles times 5 points minimum for mastery). It would also be possible not to have mastered more than half of the principles and still to have a score as high as 46 out of the 54. With the addition of anywhere from zero to eighteen points on the three item groups combining principles, the minimum score possible with mastery of more than half the principles would be 25, i.e. 25 plus 0, and the maximum score without mastery of more than half of the principles would be 64, i.e. 46 plus 18. This forty-point range does not provide much guidance in setting a cut-off score for inclusion in the study. It would, of course, have been possible simply to count the principles mastered, but it was not desired to insist that the instruction actually bring performance up from non-mastery to the mastery criterion. It was thought sufficient that the instruction produce improvement in reasoning without regard to the acquisition of newly mastered principles.

In actual fact the pre-test scores of the potential participants were distributed as in Table 3.

Table 3

Pre-Test Scores on Cornell Class Reasoning Test for Potential Subjects in College Sample

Class Reasoning Pre-test Score	Discrepancy Group Vq	Discrepancy Group Qv	Total
70-72	3	3	6
67-69	4	7	11
64-66	2	6	8
61-63	3	5	8
58-60	3	1	4
55-57	2	1	3
52-54	3	0	3
49-51	0	0	0
N	20	23	43
Mean Score	62.20	65.35	63.88
SD of Scores	5.82	3.73	4.04

In view of the relatively high scores and the small number volunteers, it was decided to include everyone who could improve by as few as four points, that is, had a score below 69. This was far from desirable, but a lower cut-off point would have reduced the total number to an unacceptable level. The sixteen students in each group with the lowest pre-test scores were asked to take the conditional reasoning pre-test and were assigned randomly to the two forms of the instructional program. Fifteen in each group completed their programs. Their mean scores were as follows:

Vq N	Vq Mean	Vq SD	Qv N	Qv Mean	Qv SD	Total N	Total Mean	Total SD
15	60.67	4.66	15	63.87	3.32	30	62.27	4.35

The difference of 3.20 between the Vq and Qv groups has a Standard Error of 1.47, producing a t value of 2.18, $p < .05$.

The conditional reasoning pre-test was the Cornell Conditional Reasoning Test (Form X), also developed by Ennis and others (1965) and following precisely the same format as the class reasoning test, but testing twelve principles of conditional reasoning instead of eight class reasoning principles. The reported reliabilities of the two tests are .83 for class reasoning and .75 for conditional reasoning (Ennis, 1965, p.IV 20-21).

The distribution of conditional reasoning pre-test scores for the students who actually participated in the study was as follows (through a clerical error one subject was given the wrong test and did not, therefore, have a pre-test score in conditional reasoning):

Table 4

Distribution of Pre-Test Scores on Cornell Conditional Reasoning Test for College Sample

Conditional Reasoning Pre-Test Score	Discrepancy Group Vq	Qv	Total
70-72	0	1	1
67-69	0	0	0
64-68	1	3	4
61-63	1	3	4
58-60	2	2	4
55-57	4	4	8
52-54	2	2	4
49-51	3	0	3
46-48	1	0	1
N	14	15	29
Mean Score	55.29	60.47	57.97
SD of Scores	4.84	4.99	5.56

As in the class reasoning test, the mean score of the Qv group is significantly higher than that of the Vq group. The difference of 5.18 has a standard error of 1.84, yielding a *t* value of 2.81, $p < .01$.

Based on the nine single principle item groups and using Ennis' criteria for mastery, the extent to which the eight principles were mastered on the class reasoning pre-test is given in Table 5.

Table 5

Number of Students in College Sample Mastering Various Numbers of Class Reasoning Principles on Pre-Test

	Number of Principles Mastered		Number of Subjects Mastering
	8		8
	7		7
	6		8
	5		4
	4		3
Mean	6.43	N	30

Instructional Programs

Upon completion of both pre-tests subjects were given the program version to which they had been assigned and requested to respond overtly to each frame, to complete the program as quickly as possible, and to return it when finished, at which time they were to take the post-tests.

The programs were developed by the investigators themselves. The verbal (VS) version was constructed first in a linear format The first section dealt with Class Reasoning Principle 1 and some basic concepts and terminology. Succeeding sections dealt with the remaining seven principles and the process of combining principles. After each principle had been explained, practice frames were provided following the format used in the tests.

A preliminary version of the VS form was tried out by a group of fourteen graduate students, who were asked to record the total time required to complete it and to criticize the wording of frames. An analysis of their errors indicated points at which changes or additional frames were needed The first draft was extensively revised on the basis of this try-out. In addition,it was learned that the program required approximately three hours to complete and that some learning resulted. The pre- and post-test scores on the Cornell Class Reasoning Test for the ten members of the try-out group who took both tests are shown in Table 6.

Table 6

Pre- and Post-test Scores of Try-out Group on Cornell Class Reasoning Test

Score	Pre-test	Post-test
64-66	0	1
61-63	1	1
58-60	0	2
55-57	1	1
52-54	2	2
49-51	1	2
46-48	4	1
43-45	1	0
Total	10	10
Mean Score	50.6	55.7
SD of Scores	5.60	6.15

The mean gain of 5.1 points had a standard error of 1.41, producing a t of 3.62, which is significant at the .01 level. This mean gain was 23.8 percent of the possible gain. It was hoped that revision of the program would produce even greater gains.

The SY form was constructed to parallel the VS form as closely as possible. Whenever possible, letter symbols were substituted for the scientific terms used in the verbal version.

Since much verbal material necessarily remained, Euler circle diagrams were inserted at various points, without explanation. One kind of problem could not be converted to symbolic form. This was the problem of being misled by empirical falsity in instances of logical validity. Hence, the SY version contained a few frames with verbal examples.

The programs were reproduced, with one frame per page and with the feedback information at the top of the following page. Table 7 shows the number of frames per section in the two versions.

Table 7

Number of Frames in Two Versions of Instructional Program

Section	Frames in VS	Frames in SY
1	63	70
2	33	34
3	20	21
4	18	18
5	16	16
6	13	13
7	9	9
8	17	17
9	24	24
Total	213	222

Sample frames from each program are included in Appendix A.

Post-testing

Post-tests were administered individually as soon as each subject completed his program. Unfortunately, a student strike and accompanying campus unrest coincided with this phase of the study. A number of the subjects went home. Most of these were persuaded to complete the program, mail it back, and obtain and return the post-test by mail. Two subjects failed to complete this process, reducing the total N of the study and unbalancing the design.

The same two tests which were used as pre-tests served as post-tests. The nature of the items, the length of the tests, the elapse of time, and the fact that the students were not informed of the correctness of their pre-test responses were judged sufficient bases for assuming negligible memory and practice effects. In any event, whatever such effects existed were assumed to be random among the four aptitude-treatment groups. The post-test scores are reported and discussed in the next chapter on "Results".

Middle School Experiment

Because the pre-test scores of the college group were so high, it was decided not to replicate the experiment at the college level, but to seek instead a group of younger subjects on whom discrepancy data were available and who would be willing to participate. Published data (Ennis and Paulus, 1965, p. V-16) indicated that mean scores between 44.7 and 53.4 could be expected between the eighth and tenth grades. Since most schools , in New York State at least, do not give scholastic aptitude tests yielding V and Q scores until the eighth or ninth grade, it appeared that a ninth grade group would be the youngest for which the requisite data would be available. Arrangements were made with the Colonie (N.Y.) Central School District to conduct the experiment at the Lishakill Middle School. Two guidance counselors agreed to make available students' scores on the California Test of Mental Maturi y, to administer the pre- and post-tests, and to manage the distribution and collection of the programs. The investigators set the criteria for inclusion and assigned selected students randomly to program versions.

The procedures with this phase of the study were the same as those described for the college sample with the following exceptions:

1. The programs were slightly modified to simplify wording somewhat. Essentially, the frames remained unchanged.
2. The conditional reasoning test to study transfer effects was not used.
3. Parental permission was secured for participation and granted in all instances.
4. The students did not receive a stipend for participating.
5. A total of forty subjects was selected, with two in each of the four aptitude-treatment groups designated as alternates. Their scores would be used only if any of the regular subjects failed to complete all aspects, in order to obtain eight cases in each of the four cells. The alternates had somewhat smaller aptitude discrepancies than the regularly selected subjects.

The criteria for selection of the two groups were the following scores on the CTMM:

Vq: V score > 114 and V-Q > 12
Qv: Q score > 114 and V-Q > 12

The median verbal and quantitative aptitude scores of the 31 ninth grade participants are shown in Table 8. Because they were chosen from a group of several hundred rather than several thousand, they are neither as discrepant as the college group, nor do the two groups differ as markedly. The lowest verbal score of the Vq group is only three points lower than the highest verbal score of Qv group. Similarly, the lowest quantitative score of the Qv group is only five points higher than the highest quantitative score of the Vq group. As in the college sample, the Vq-VS group happens to be somewhat less discrepant than the other three groups.

Table 8

Verbal and Quantitative Aptitude Scores of Ninth Grade Students Participating in Study

		Verbal Aptitude		Quantitative Aptitude		Discrepancy	
Group	N	Median	Range	Median	Range	Median	Range
Vq-VS	7	122.0	115-123	105.0	93-110	15.0	13-27
Vq-SY	8	118.5	116-132	100.5	89-109	21.5	13-29
Total Vq	15	120.0	115-132	103.0	89-110	21.0	13-29
Qv-VS	8	102.5	87-112	119.5	116-132	19.5	16-31
Qv-SY	8	98.5	93-107	118.5	115-129	20.5	14-33
Total Qv	16	100.5	87-112	119.0	115-132	20.0	14-33

The class reasoning pre-test scores of the 31 ninth grade subjects who completed the program are given in Table 9.

Table 9

Distribution of Class Reasoning Pre-test Scores for Ninth Grade Samples

Pre-test Score	Discrepancy Group Vq	Qv	Total
61-63	1	2	3
58-60	1	3	4
55-57	0	2	2
52-54	2	2	4
49-51	4	1	5
46-48	1	2	3
43-45	3	3	6
40-42	2	0	2
37-39	0	0	0
33-36	1	1	2
Total	15	16	31
Mean Score	47.67	52.25	50.03
SD of Scores	6.71	7.50	7.49

The overall mean of 50.03 falls between the eighth and tenth grade values reported by Ennis and compares with the mean of 62.27 for the college sample. This difference of 12.24 is obviously highly significant statistically. In both samples, the Qv group had a higher mean score than the Vq group, but while the difference in the college sample was significant at the .05 level, the difference of 4.58 in the ninth grade sample has a standard error of 2.55, yielding a t value of 1.80 which has a probability in excess of .05.

College Norming Phase

To obtain a broader view of college students' deductive reasoning abilities, a sample of 200 names (50 from each class) was drawn systematically from alphabetized registration lists at SUNY-Albany in the fall of 1970. Each person whose name was drawn was offered two dollars to take the Cornell Class Reasoning Test at a group testing session or individually at an arranged time. The selected students were offered an additional dollar if they would also take the Cornell Conditional Ressoning Test. When both tests were taken, the order was reversed for every other person.

Of the 200 students invited, a few had already withdrawn from college and some declined to participate. A total of 108 appeared to take the class reasoning test and 79 of these plus two other subjects took the conditional reasoning test. The examinees were distributed by class and sex, as shown in Table 10:

Table 10

Number of College Students in Norming Sample by Class and Sex

Class	Class Reasoning Test			Conditional Reasoning Test		
	M	F	T	M	F	T
1971	9	21	30	8	15	23
1972	11	17	28	9	12	21
1973	10	17	27	7	13	20
1974	10	13	23	6	11	17
Total	40	68	108	30	51	81

V

Analysis of Findings

The summaries and analyses of data in this section are based on the various scores for each participant in the study, which are reported in full in the appendix C. In the discussion that follows, the results of the testing of the broad sample of college students are presented first to provide a basis for interpreting the scores of the experimental groups. Then the means of the various sub-groups in the two experimental situations are presented and differences examined. Finally, various statistical analyses are employed in an effort to explain some of the features of the results.

Deductive Reasoning Norms in College

A systematic sample of 108 college students, stratified by class, was administered the Cornell Class Reasoning Test (Form X) at the beginning of the 1970-71 academic year. The Cornell Conditional Reasoning Test (Form X) was taken by 81 students, including 79 of the 108 just mentioned. The mean scores on both tests are shown in the first line of Table 11. These means are based on raw scores representing number of correct items out of a possible 72. The mean for class reasoning of 60.82 is equivalent to a total score of 83.23 by the scoring system used by Ennis, $(R - W/2) + 27$. The mean for conditional reasoning of 53.89 is equivalent to Ennis' total score of 71.84. The class reasoning score (83.23) is ten points higher than that reported by Ennis for twelfth grade students (73.4) and the conditional reasoning score (71.84) is fifteen points higher than that of Ennis' eleventh graders (56.6). (Ennis and Paulus, 1965, p. V-16).

The raw score means for males and females were within two points of each other on both tests. The t values given in Table 11 reveal that the slight differences in scores between the sexes were not statistically significant.

Table 11

Mean Scores on Cornell Class Reasoning and Conditional Reasoning Tests for College Norming Sample by Sex

	Class Reasoning Test				Conditional Reasoning Test			
Group	N	Mean	SD	SE_M	N	Mean	SD	SE_M
Total	108	60.82	6.74	0.65	81	53.89	6.78	0.75
Male	40	59.70	6.94	0.91	30	54.47	6.61	1.21
Female	68	61.49	6.53	0.79	51	53.55	6.78	0.96
Difference	--	1.79	--	1.20	--	0.92	--	1.53
t			1.49				0.60	

The simple median scores for each of the four college classes on the two tests are given in Table 12. The freshmen scored slightly above the total median on both tests, while the seniors scored slightly below it. There is not, however, a pattern of decline through the college years, since the juniors scored highest on both tests.

Table 12

Median Scores on Cornell Class Reasoning and Conditional Reasoning Tests for College Samples, by College Class

College Class	Class Reasoning N	Class Reasoning Mdn	Conditional Reasoning N	Conditional Reasoning Mdn
1971	30	59.5	23	51.0
1972	28	64.5	21	56.0
1973	27	59.0	20	53.5
1974	23	62.0	17	53.0
Total	108	61.0	81	53.0

To test whether significant differences were present among the classes, the scores for each class were dichotomized into those exceeding and not exceeding the total median. The two by four tables for the two reasoning tests are combined in Table 13. Differences as large as those among the classes on the conditional reasoning test may be expected by chance more than fifty percent of the time. Differences as large as those on the class reasoning test, however, may be expected less than five percent of the time and are therefore statistically significant. To test further whether a pattern existed across the college years, the data were recast into a fourfold table comparing the first two years with the last two years. The resulting Chi square of 0.121 with one degree of freedom had an associated probability in excess of .70, indicating that the previous significant Chi square for class reasoning probably represented a departure on the part of the junior class (1972) alone, rather than any pattern across the years.

Table 13

Number of Scores on Class Reasoning and Conditional Reasoning Tests for Each College Class Exceeding and Not Exceeding Overall Median

Class	Class Reasoning Exceeding Median	Class Reasoning Not Exceeding Median	Conditional Reasoning Exceeding Median	Conditional Reasoning Not Exceeding Median
1971	9	21	9	14
1972	18	10	13	8
1973	12	15	10	10
1974	13	10	8	9
Total	52	56	40	41
Chi square	9.12		2.26	
p (df - 3)	$<.05$		$>.50$	

It will be noted from Tables 11 and 12 that for class reasoning both the mean and the median were 61, and for conditional reasoning the mean was 54 and the median, 53. This similarity between the means and medians suggests that the scores for college students were distributed symetrically and probably normally, although a test for normality was not applied. Because there were no significant sex differences or effects of studying logic and no systematic differences across the college years, a single set of tentative college norms was drawn up in the form of a percentile distribution, rounded to whole scores. The scores on each test associated with each decile and the top and bottom percentiles are given in Table 14.

Table 14

Percentile Scores for Sample of College Students on Cornell Tests of Class and Conditional Reasoning

Percentile	Total Score: Class Reasoning (N=108)	Total Score: Conditional Reasoning (N=81)
99	72	70
90	69	63
80	67	60
70	66	57
60	64	55
50	61	53
40	59	51
30	57	50
20	54	49
10	52	45
1	43	37

An indication of how closely the distribution approximates the normal one can be gained by noting the score values of the 80th and 20th percentiles. In a normal distribution these should be .84 SD above and below the mean, respectively. This would be about six points in these distributions. The 80th percentile in both distributions (67 and 60) is in fact six points higher than the mean (61 and 54). The 20th percentile (54 and 49) is seven points below the mean for class reasoning and five points below for conditional reasoning. Both of these values are within a point of what would be predicted, but suggest that the conditional reasoning distribution may be skewed more toward the lower end than the one for class reasoning.

Since there were in the sample 79 students who took both the class reasoning and the conditional reasoning test, it was possible to determine that the Pearson coefficient of correlation between the two tests was .745. Since the order in which the tests were taken might affect the means, as well as the correlation between them, the order was randomized, 38 students taking the class reasoning test first and 41 taking it after having taken the conditional reasoning test. The means and standard deviations for the two orders and for the total group of 79 taking both tests are given in Table 15, along with the corresponding data for all subjects taking each of the tests. Oddly enough the slight differences with respect to order favor the test taken first. Apparently a fatigue effect was greater than any practice effect. The differences, however, are not significant, the *t* values, 0.59 and 1.31 having associated probabilities of about .55 and .20, respectively.

Table 15

Mean Scores on Class Reasoning and Conditional Reasoning Tests for College Students Taking Both Tests, by Test Order

Group	Class Reasoning				Conditional Reasoning			
	N	Mean	SD	SE_M	N	Mean	SD	SE_M
Total Taking Each Test	108	60.82	6.74	0.65	81	53.89	6.78	0.75
Total Taking Both Tests	79	60.11	6.85	0.77	79	53.94	6.34	0.72
Class First	38	60.58	7.37	1.19	38	52.94	7.00	1.13
Conditional First	41	59.66	6.52	1.02	41	54.85	5.79	0.90
Difference	--	0.92	---	1.57	--	1.91	--	1.45
t			0.59				1.31	

About one fifth of the students in the sample reported that they either had taken or were taking one or more courses dealing with formal or symbolic logic. Table 16 indicates that those who had had some logic instruction scored only slightly higher on the average than those who had never studied logic. For neither type of reasoning was the difference statistically significant.

Table 16

Mean Scores on Class and Conditional Reasoning Tests For College Students Who Had and Had Not Studied Logic

Group	Class Reasoning				Conditional Reasoning			
	N	Mean	SD	SE_M	N	Mean	SD	SE_M
Studied Logic	23	61.04	8.05	1.68	20	55.75	8.03	1.80
No Logic Study	85	60.76	6.34	0.69	61	53.28	6.20	0.79
Difference	--	0.28	--	1.82	--	2.47	--	1.96
t			0 15				1.26	

Grade point averages were reported by 54 students, all of whom took the class reasoning test and 40 of whom took the conditional reasoning test. These students, with one exception, were juniors and seniors, since letter grades are not given to freshmen and sophomores. The median reasoning scores of students whose grade point averages fell within certain intervals are given in Table 17. The overall median scores of this sub-sample are almost exactly the same as those for all juniors and seniors in the total sample, of whom the sub-sample included 90 percent. The relationship between grade point average and conditional reasoning score is essentially linear, with a product-moment correlation of .359 The distribution of class reasoning scores, however, appears to be bimodal, the relationship being much stronger at the upper and lower GPA range than in the middle of it. The overall Pearson r was .314. It should be noted that few measures, other than high school average, correlate as high as .30 with college grade point average.

Table 17

Median Scores on Class and Conditional Reasoning Tests for College Students by Grade Point Average

Grade Point Average	Class Reasoning		Conditional Reasoning	
	N	Mdn	N	Mdn
Above 3.4	5	69.0	2	63.0
3.0 - 3.3	14	61.0	9	56.0
2.6 - 2.9	24	64.0	20	53.0
Below 2.6	11	55.0	9	50.0
Total	54	61.0	40	53.5
r		.314		.359

Experimental Results

The experimental data consist of pre-test and post-test scores on the Cornell Class Reasoning Test for the college sample and the ninth grade sample and on the Cornell Conditional Reasoning Test for the college sample only. Pre- and post-test data on the number of class reasoning principles mastered are also available. All data are reported here by aptitude discrepancy type (Vq or Qv), by treatment (VS or SY program), by the two combined (Vq-VS, Vq-SY, Qv-VS, or Qv-SY), and by matching of type and treatment (Vq-VS and Qv-SY considered matched, Vq-SY and Qv-VS considered mismatched). Direct learning of class reasoning is discussed first, after which transfer to conditional reasoning is considered.

Direct learning of class reasoning. The mean scores in class reasoning before and after instruction are given in Tables 18 and 19 for various sub-groups in each sample. The programs produced an overall gain of 2.30 with the college sample, which is significant at the .05 level. With the ninth grade sample, however, the overall gain of 1.68 was not statistically significant. None of the sub-groups in the ninth grade made a significant gain, although that of the "matched" groups exceeded a probability of .05 only slightly. In the college sample, on the other hand, the gains of the Qv discrepancy group and of students assigned the SY program were significant. These gains were largely attributable to the particular effectiveness of the SY program with the Qv group, which also led to a significant gain for the "matched" groups, even though the VS program was of less than average effectiveness with the Vq group.

Table 18

Mean Scores on Class Reasoning Test for College Sample Before and After Instruction

Group	N	Before Mean	SD	After Mean	SD	Gain Mean	SE	t
Total	30	62.27	4.35	64.57	5.19	2.30	0.74	3.11**
Discrepancy Type								
Vq	15	60.67	4.66	62.93	5.56	2.26	1.19	1.91
Qv	15	63.87	3.32	66.20	4.18	2.33	0.98	2.38*
Treatment								
VS	15	62.40	4.03	63.47	5.57	1.07	0.85	1.26
SY	15	62.13	4.65	65.67	4.50	3.54	1.19	2.97**
Experimental Condition								
Vq-VS	8	60.50	4.03	62.13	6.01	1.63	1.09	1.50
Vq-SY	7	60.86	5.28	63.86	4.82	3.00	2.49	1.20
Qv-VS	7	64.57	2.72	65.00	4.44	0.43	1.49	0.29
Qv-SY	8	63.25	3.67	67.25	3.49	4.00	1.11	3.60**
Type-Treatment Match								
Matched	16	61.88	4.09	64.69	5.42	2.81	0.79	3.56**
Mismatched	14	62.71	4.59	64.43	4.81	1.72	1.36	1.26

* $p < .05$

** $p < .01$

Table 19

Mean Scores on Class Reasoning Test for Ninth Grade Sample Before and After Instruction

Group	N	Before Mean	Before SD	After Mean	After SD	Gain Mean	Gain SE	t[1]
Total	31	50.03	7.49	51.71	7.45	1.68	1.10	1.53
Discrepancy Type								
Vq	15	46.67	6.71	50.00	8.35	2.33	1.88	1.24
Qv	16	52.25	7.50	53.31	6.07	1.06	1.32	0.80
Treatment								
VS	15	50.60	7.82	52.87	6.26	2.27	1.62	1.40
SY	16	49.50	7.12	50.63	8.27	1.13	1.76	0.64
Experimental Condition								
Vq-VS	7	44.57	5.42	49.43	5.65	4.86	2.79	1.74
Vq-SY	8	50.38	6.56	50.50	10.11	0.12	2.58	0.05
Qv-VS	8	55.88	5.42	55.88	5.11	0.00	1.73	0.00
Qv-SY	8	48.63	7.53	50.75	5.87	2.12	2.14	0.99
Type-Treatment Match								
Matched	15	46.73	6.93	50.13	5.80	3.40	1.65	2.06
Mismatched	16	53.13	6.61	53.19	8.47	0.06	1.57	0.04

[1]None of *t* values is significant at .05 level.

Although the ninth graders made smaller average gains than the college students, approximately the same proportion of them made some gain (see Table 20). In both samples the percent of each discrepancy type and the percent of those assigned each program version who gained were approximately the same and the actual proportions were similar in the two samples. For the ninth grade sample approximately the same percent of those whose program version matched their discrepancy type made gains as of those who were mismatched, but this was not the case in the college sample, where a much larger proportion of matched students gained than of those who were mismatched. Thus, there is a suggestion of interaction between treatment and discrepancy type in the college sample.

Table 20

Number and Percent of Subjects in College and Ninth Grade Samples Whose Score on Class Reasoning Test Increased Following Instruction

	College (N=30)		Ninth Grade (N=31)	
Group	N	%	N	%
Total	21	70.0	21	67.7
Discrepancy Type				
Vq	11	73.3	11	73.3
Qv	10	66.7	10	62.5
Treatment				
VS	10	66.7	11	73.3
SY	11	73.3	10	62.5
Experimental Condition				
Vq-VS	7	87.5	6	85.7
Vq-SY	4	57.1	5	62.5
Qv-VS	3	42.9	5	62.5
Qv-SY	7	87.5	5	62.5
Type-Treatment Match				
Matched (Vq-VS, Qv-SY)	14	87.5	11	73.3
Mismatched (Vq-SY, Qv-VS)	7	50.0	10	62.5

Table 21 shows the number of college students in each of the four experimental sub-groups who demonstrated mastery of each of the eight class reasoning principles before and after completing the instructional program. Each group gained slightly in the number of principles mastered, the mean being 0.6 principles. Gains were made for all except two principles, the first (for which no gain was possible) and the seventh. The greatest gains were on Principles 4 and 8. Comparable data for the ninth grade are given in Table 22. The mean gain here is 0.5 principles and one sub-group (Qv-VS) declined slightly. Losses occurred for three principles, the first, fifth, and eighth. The ninth graders' greatest gains occurred on Principles 2 and 3. Thus, the program seems to have been effective in different ways at the two educational levels.

Table 21

Number and Percent of College Sample Demonstrating Mastery of Each of Eight Principles of Class Reasoning Before and After Instruction

		Group				Total	
Principle		Vq-VS N=8	Vq-SY N=7	Qv-VS N=7	Qv-SY N=8	No. N=30	%
1	Pre	8	7	7	8	30	100
	Post	8	7	7	8	30	100
2	Pre	6	7	6	8	27	90
	Post	8	6	7	8	29	97
3	Pre	4	5	6	6	21	70
	Post	5	6	5	8	24	80
4	Pre	4	6	5	8	23	77
	Post	8	6	7	8	29	97
5	Pre	8	6	6	7	27	90
	Post	7	7	7	8	29	97
6	Pre	4	4	6	7	21	70
	Post	3	5	7	7	22	73
7	Pre	8	5	7	7	27	90
	Post	7	7	7	6	27	90
8	Pre	3	4	4	6	17	57
	Post	6	5	5	7	23	77
Total	Pre	45	44	47	57	193	--
	Post	52	49	52	60	213	--
Mean Principles	Pre	5.6	6.3	6.7	7.1	6.5	81
Mastered	Post	6.5	7.0	7.4	7.5	7.1	89

Table 22

Number and Percent of Ninth Grade Sample Demonstrating Mastery of Each of Eight Principles of Class Reasoning Before and After Instruction

Principle		Group Vq-VS N=7	Vq-SY N=8	Qv-VS N=8	Qv-SY N=8	Total No. N=31	%
1	Pre	7	8	8	7	30	97
	Post	6	7	8	8	29	94
2	Pre	4	3	8	5	20	64
	Post	6	6	8	8	28	90
3	Pre	0	0	2	0	2	06
	Post	1	2	2	3	8	26
4	Pre	3	6	7	4	20	64
	Post	3	6	6	6	21	67
5	Pre	7	7	8	6	28	90
	Post	5	7	8	6	26	84
6	Pre	0	1	1	0	2	06
	Post	2	0	1	0	3	10
7	Pre	1	3	5	1	10	32
	Post	3	3	4	3	13	42
8	Pre	1	5	4	4	14	45
	Post	1	4	5	3	13	42
Total	Pre	23	33	43	27	126	--
	Post	27	35	42	37	141	--
Mean Principles Mastered	Pre	3.3	4.3	5.4	3.4	4.1	51
	Post	3.9	4.8	5.3	4.6	4.6	57

These differential effects apply not only to particular principles but to sub-groups in the samples. Whereas both for total score and for number showing gains, the matching of discrepancy type and treatment seemed to have had some effect only in the college sample, it can be seen in Table 23 that with regard to the mastery of principles, the interaction effect is noticeable only in the ninth grade sample.

Table 23

Mean Number of Class Reasoning Principles Mastered Before and After Instruction by Sub-Groups of College and Ninth Grade Samples

	College			Ninth Grade		
Group	N	Pre	Post	N	Pre	Post
Total	30	6.5	7.1	31	4.1	4.6
Type						
Vq	15	5.9	6.7	15	3.7	4.1
Qv	15	6.9	7.5	16	4.4	4.9
Treatment						
VS	15	6.1	6.9	15	4.4	4.6
SY	15	6.7	7.3	16	3.8	4.5
Type-Treatment						
Matched	16	6.4	7.0	15	3.1	4.3
Mismatched	14	6.5	7.2	16	4.8	4.8

Transfer to conditional reasoning. The college sample was also given a test in conditional reasoning before and after instruction in class reasoning. The scores for various sub-groups are presented in Table 24. Only the Vq discrepancy group made a statistically significant transfer from class reasoning instruction to conditional reasoning performance, largely when taught by the VS version. By contrast, the Qv group when taught by the VS version, showed a slight loss in conditional reasoning. Although the gain made by the group taught by a program version matched to its discrepancy type was not statistically significant, it was over eight times as great as that of the mismatched group.

Table 24

Mean Scores on Conditional Reasoning Test for College Sample Before and After Instruction

Group	N	Before Mean	SD	After Mean	SD	Gain Mean	SE	t
Total	29	57.97	5.56	59.14	5.22	1.17	0.70	1.67
Discrepancy Type								
Vq	14	55.29	4.84	57.14	4.16	1.85	0.83	2.24*
Qv	15	60.47	4.99	61.00	5.42	0.53	1.21	0.44
Treatment								
VS	15	55.87	5.11	57.13	4.87	1.26	0.85	1.49
SY	14	60.21	5.13	61.29	4.70	1.08	1.27	0.84
Experimental Condition								
Vq-VS	8	53.75	4.82	56.38	4.41	2.63	1.28	2.05
Vq-SY	6	57.33	4.07	58.17	3.53	0.84	1.00	0.84
Qv-VS	7	58.29	4.30	58.00	5.21	-0.29	0.90	0.32
Qv-SY	8	62.38	4.77	63.63	4.06	1.25	2.25	0.56
Type-Treatment Match								
Matched	16	58.06	6.45	60.00	5.58	1.94	1.22	1.59
Mismatched	13	57.85	4.22	58.08	4.51	0.23	0.64	0.36

* $p < .05$

Statistical Analyses

The statistical tests presented to this point have indicated no significant differences between groups with treatment matched to discrepancy type and those which were mismatched, whether in total score, number showing score increase, or number achieving mastery of a principle. Yet in nearly all instances, the directionality of observed differences has been that hypothesized.

Analyses of covariance. Since the correlation between pre-test and post-test scores is high (.75), since the randomly assigned ninth grade groups differed considerably in pre-test score, and since the chief interest of the study is whether a treatment-discrepancy type interaction exists, an analysis of covariance was carried out for the three sets of data, with the pre-test score as covariate. If a significant interaction were shown by this more sensitive parametric test, then it would be necessary to determine whether all of the assumptions for such tests were met. However, since some of the assumptions are met and since other tests indicated a lack of significant interaction, such determination was not made prior to performing the covariance analyses, the results of which are presented in Tables 25, 26, and 27. They show that none of the F values, for discrepancy type, program version, or interaction between the two was significant at the .05 level.

Table 25

Summary of Analysis of Covariance of Scores
of College Sample on Class Reasoning Test
(N=28)

Source	df	SSx	SSy	SSxy	$\frac{(SSxy)^2}{SSx}$	SSy[1]	ms	F	p
Error	23	469.43	675.43	411.00	359.84	315.59	13.72	--	--
Interaction	1	2.89	1.57	-1.28	--	5.82	5.82	0.42	N.S.[1]
Subtotal 1	--	472.32	677.00	409.82	355.59	321.41	--	--	--
Treatments	1	4.32	17.29	-8.65	--	35.42	35.42	2.58	N.S.
Subtotal 2	--	473.75	692.72	402.35	341.71	351.01	--	--	--
Discrepancy Types	1	60.04	51.57	55.64	--	0.14	0.14	0.01	N.S.
Subtotal 3	--	529.47	727.00	466.64	411.27	315.73			

X (covariate) = pre-test score
Y = post-test score
[1] N.S. = not significant at .05 level

Table 26

Summary of Analysis of Covariance of Scores of Ninth Grade Sample on Class Reasoning Test (N=32)

Source	df	SSx	SSy	SSxy	$\frac{(SSxy)^2}{SSx}$	SSy^1	ms	F	p
Error	27	1264.12	1536.25	881.38	614.52	921.73	34.14	--	--
Interaction	1	305.72	84.50	160.87	--	6.05	6.05	0.18	N.S.[1]
Subtotal 1	--	1569.84	1620.75	1042.25	691.97	928.78	--	--	--
Treatments	1	9.60	28.12	15.94	--	7.67	7.67	0.22	N.S.
Subtotal 2	--	1273.72	1565.37	899.32	634.97	930.40	--	--	--
Discrepancy Types	1	157.53	98.00	124.25	--	1.17	1.17	0.03	N.S.
Subtotal 3	--	1421.65	1634.25	1005.63	711.35	922.90	--	--	--

X (covariate) = pre-test score
Y = post test score
[1] N.S. = not significant at .05 level

Table 27

Summary of Analysis of Covariance of Scores of College Sample on Conditional Reasoning Test (N=24)

Source	df	SSx	SSy	SSxy	$\frac{(SSxy)^2}{SSx}$	SSy^1	ms	F	p
Error	19	468.33	422.50	266.67	151.84	270.66	14.25	--	--
Interaction	1	.68	5.04	35.75	--	36.37	36.37	2.55	N.S.
Subtotal 1	--	469.01	427.54	302.42	193.25	234.29	--	--	--
Treatments	1	48.16	30.38	38.25	--	37.15	37.15	2.61	N.S.
Subtotal 2	--	517.17	457.92	340.67	224.41	233.51	--	--	--
Discrepancy Types	1	112.66	51.04	75.84	--	37.14	37.14	2.61	N.S.
Subtotal 3	--	629.83	508.96	416.51	275.44	233.52	--	--	--

X (covariate) = pre-test score
Y = post-test score
1/ N.S. = not significant at .05 level

Symbolic component sub-scores. Each item group in the Cornell Class Reasoning Test, whether testing a single principle or a combination of them, was comprised of three types of items, designated CF, SY, and SU. In each item group there are four items of the CF (concrete familiar) type, meaning that the "... content mentioned is concrete articles and qualities with which the subject has been associated." (Ennis and Paulus, 1965, p.IV-8). One item in each group is of the SU (suggestive)type in which the truth status of the content in the conclusion is known to the subject and is at variance with the validity status of the argument. It was this type that necessitated the inclusion of some verbal content in the SY program used in this study. The sixth test item in each group is of the SY (symbolic) type, using letters instead of words as content in the statements, as was the case with the majority of frames in the SY and of only a few in the program VS version.

Because only one-sixth of the test uses a symbolic format, while five-sixths of the items use words as content, the possibility was envisioned that the program versions might have differential practice effects upon test item types, rather than or in addition to, learning and transfer effects relative to principles of deductive reasoning. The twelve test items making up the SY component were therefore scored separately to provide an SY subscore for each subject. The various sub-group means for these sub-scores are given for the college sample in Table 28 and for the ninth grade sample in Table 29.

The overall gain by both samples on these items was statistically significant, but in both samples the gain is largely attributable to that made by the Vq group using the SY program. This group may not have been proficient at using symbols in reasoning and the SY program may have taught these students the use of symbols as well as class reasoning principles.

Table 28

Mean Sub-scores on the Symbolic (SY) Component
of the Class Reasoning Test for College Sample
Before and After Instruction

Group	N	Before Mean	Before SD	After Mean	After SD	Mean Gain	SE	t
Total	30	8.9	1.96	10.2	1.70	1.3	0.38	.34 **
Discrepancy Type								
Vq	15	7.9	1.95	9.6	1.62	1.7	0.63	2.65
Qv	15	9.9	1.41	10.7	1.57	0.8	0.48	1.81
Treatment								
VS	15	9.1	1.63	9.5	1.67	0.4	0.42	0.93
SY	15	8.7	2.21	10.8	1.22	2.1	0.55	3.88 **
Experimental Condition								
Vq-VS	8	8.4	1.65	9.0	1.11	0.6	0.43	1.47
Vq-SY	7	7.4	2.11	10.3	1.83	2.9	0.93	3.08 *
Qv-VS	7	10.0	1.07	10.1	1.94	0.1	0.74	0.19
Qv-SY	8	9.8	1.64	11.3	0.83	1.5	0.53	2.83 *
Type Treatment Match								
Matched (Vq-VS, Qv-SY)	16	9.1	1.78	10.1	1.62	1.0	0.36	2.95 **
Mismatched (Vq-SY, Qv-VS)	14	8.7	3.10	10.2	1.90	1.5	0.70	2.14

* Significant at .05 level
** Significant at .01 level

Table 29

Mean Sub-scores on the Symbolic (SY) Component of the Class Reasoning Test for Ninth Grade Sample Before and After Instruction

Group	N	Before Mean	Before SD	After Mean	After SD	Mean Gain	SE	t
Total	31	6.4	1.58	7.1	1.48	0.7	.38	2.13 *
Discrepancy Type								
Vq	15	6.0	1.59	7.3	1.09	1.3	.40	3.30 **
Qv	16	6.9	1.43	6.9	1.30	0.0	.43	0.14
Treatment								
VS	15	6.8	1.56	7.1	1.22	0.3	.45	0.73
SY	16	6.1	1.54	7.1	1.41	1.0	.43	2.33 *
Experimental Condition								
Vq-VS	7	5.7	1.58	6.9	0.64	1.2	1.44	2.20
Vq-SY	8	6.3	1.56	7.8	1.56	1.5	.59	2.54 *
Qv-VS	8	7.8	0.66	7.4	1.49	-0.4	.61	0.62
Qv-SY	8	6.0	1.50	6.5	0.87	0.5	.59	0.85
Type-Treatment Match								
Matched (Vq-VS, Qv-SY)	15	5.9	1.54	6.7	0.79	0.8	.40	2.00
Mismatched (Vq-SY, Qv-VS)	16	7.0	1.41	7.6	1.54	0.6	.48	1.16

* significant at .05 level
** significant at .01 level

The correlation between pre-test and post-test scores on these twelve SY items was much lower than that for the class reasoning test as a whole. Whereas for the college sample the pre-post correlation for total scores was .721, the coefficient for the SY component was only .357. For the ninth grade sample the total score correlation was .665 and that for the SY component, .322. Of course the lower reliability of the component score will depress the correlation to some extent. However, the fact that the component correlations are only about half of those for total score again suggests the possibility that one version of the program may have taught how to handle test items involving symbols while the other did not.

Examination of the correlations for each of the experimental sub-groups, given in Table 30, does not, however, make such an effect apparent. The operation of a variety of factors, in addition to the instability of the correlations due to the small numbers in each subgroup, makes their interpretation difficult. The correlations between the total scores and the SY component scores on the class reasoning pre-test were .692 for the college sample and .538 at the ninth grade level, indicating that the SY items were measuring something in addition to reasoning ability, such as ability to deal with abstract symbols. That this effect would be more marked at the ninth grade level would appear reasonable to expect.

Table 30

Correlations Between Pre-test and Post-test Values of Total Score and Symbolic Component Score on Class Reasoning Test for Both Samples

	College		Ninth Grade	
Group	Total	SY	Total	SY
VqVS	.931	.674	.289	.527
VqSY	.638	.226	.789	.436
QvVS	.677	.366	.672	-.158
QvSY	.692	.415	.797	.108
Total	.721	.357	.665	.322

VI

Conclusions

In this concluding chapter, the evidence supporting and refuting the hypotheses proposed earlier is summarized. Some general conclusions are drawn, and recommendations are offered for educational practitioners and researchers.

Findings Relating to Direct Learning Hypotheses

The first hypothesis was that individuals with high verbal aptitude and low quantitative aptitude (Vq) would learn more about class reasoning from a verbal program (VS) than from one using symbols (SY). Three kinds of evidence are available for each of the two samples. These consist of (1) total scores before and after instruction, (2) number whose scores increased, and (3) number of principles mastered before and after instruction.

(1) Total score for Vq groups:

	College Sample			Ninth Grade Sample		
	Before	After	Gain	Before	After	Gain
VS program	60.50	62.13	1.63	44.57	49.43	4.86
SY program	60.86	63.86	3.00	50.38	50.50	0.12

The college groups were roughly equivalent at the outset and contrary to the hypothesis, those who had the SY program gained most, although the difference between the posttest means is not significant ($t = 0.62$). In the ninth grade sample the groups were initially quite different, but not significantly so ($t = 1.87$). After instruction, the gap had almost been closed, lending some support to the hypothesis.

(2) Number in Vq groups showing gain:

	College Sample	Ninth Grade Sample
VS program	7 out of 8	6 out of 7
SY program	4 out of 7	5 out of 8

In the two samples, 13 out of 15 Vq students gained with the VS program, 9 out of 15 gained with the SY. The direction supports the hypothesis, but the difference lacks statistical significance.

(3) Principles mastered by Vq groups:

	College Sample			Ninth Grade Sample		
	Before	After	Gain	Before	After	Gain
VS program	5.6	6.5	0.9	3.3	3.9	0.6
SY program	6.3	7.0	0.7	4.3	4.8	0.5

In the two samples, Vq students with the VS program achieved mastery of an average of 0.75 principles, as against 0.60 for those with the SY. Again the direction of the difference is consistent with the hypothesis, but it is not a significant one.

Thus, five of the six items of evidence indicate statistically non-significant differences consistent with the hypothesis and one, also not significant, is opposite to it. One reason why a more marked advantage was not evident for the VS program was suggested by Table 28, in which it was shown that on the symbolic (SY) test items the Vq students benefited more from the SY program, thus offsetting benefits from the VS program with respect to the learning of class reasoning.

According to the second hypothesis, Qv students were expected to gain more from the SY program than from the VS program. Again three kinds of evidence can be presented.

(1) Total score for Qv groups:

	College Sample			Ninth Grade Sample		
	Before	After	Gain	Before	After	Gain
SY program	63.25	67.25	4.00	48.63	50.75	2.12
VS program	64.57	65.00	0.43	55.88	55.88	0.00

The Qv college students assigned to the SY program scored lower than those assigned the VS version prior to instruction and higher following it. The difference between the post-test means is not, however, statistically significant ($t = 1.93$). The ninth grade groups were significantly different on the pre-test ($t = 2.20$), but were not on the post-test ($t = 1.86$). This change is consistent with the hypothesis.

(2) Number in Qv group showing gain:

	College Sample	Ninth Grade Sample
SY program	7 out of 8	5 out of 8
VS program	3 out of 7	5 out of 8

A greater proportion of the college Qv students gained with the SY program than with the VS program, but the difference is not significant (Chi square 3.49; df, 1; $p = .05 - .10$). There was no difference between the two program versions for the ninth graders.

(3) Principles mastered by Qv groups:

	College Sample			Ninth Grade Sample		
	Before	After	Gain	Before	After	Gain
SY program	7.1	7.5	0.4	3.4	4.6	1.2
VS program	6.7	7.4	0.7	5.4	5.3	-0.1

Here the ninth grade evidence is consistent with the hypothesis, but the college data are inconsistent with it. Overall, in the two samples, the Qv students achieved mastery of an average of 0.8 principles with the SY program and 0.3 with the VS program, the direction being supportive of the hypothesis.

Four of the six indicators tend to support the second hypothesis, one tends to refute it, and one does neither. Again, however, none of the evidence is statistically significant.

The third hypothesis predicted greater learning from a program matched with the student's aptitude discrepancy type than from one that was mismatched.

(1) Effect of matching on total scores:

	College Sample			Ninth Grade Sample		
	Before	After	Gain	Before	After	Gain
Matched	61.88	64.69	2.81	46.73	50.13	3.40
Mismatched	62.71	64.43	1.72	53.13	53.19	0.06

The matched college group had a slightly lower mean score at the outset and a slightly higher score after instruction than the mismatched group, tending to favor the hypothesis. The ninth grade groups differed significnntly before instruction ($t = 2.63$, $p < .05$), but did not differ significantly afterward ($t = 1.17$). While not significant, the data from both samples support the efficacy of matching.

(2) Number in matched groups showing gain:

	College Sample	Ninth Grade Sample
Matched	14 out of 16	11 out of 15
Mismatched	7 out of 14	10 out of 16

Slightly more than 80 percent of the matched students in the two samples gained through instruction, compared with about 57 percent of those who were mismatched. This difference, while supportive of the hypothesis, is not significant statistically (Chi square, 3.17; df, 1; p = 0.5 -.10).

(3) Principles mastered by matched groups:

	College Sample			Ninth Grade Sample		
	Before	After	Gain	Before	After	Gain
Matched	6.4	7.0	0.6	3.1	4.3	1.2
Mismatched	6.5	7.5	0.7	4.8	4.8	0.0

In both samples, the matched groups achieved mastery of a mean of 0.8 principles, while the mismatched mastered 0.4, The slight difference in the college sample is inconsistent with the hypothesis, whereas the substantial difference among the ninth graders is supportive of it.

With respect to matching, then, five of the six indicators are in a direction supportive of the hypothesis and one is contrary, though none is of statistical significance. Also bearing on this hypothesis is the information gained from two analyses of covariance, neither of which revealed a significant interaction effect in post-test scores between treatment and discrepancy type when pre-test score was the co-variate.

Findings Relating to Transfer Hypotheses

The three final hypotheses parallel the first three in predicting, for the same combinations of discrepancy type and program version, greater transfer from class reasoning instruction to performance on a conditional reasoning test.

Hypothesis 4 concerned the Vq group. The total score results on the conditional reasoning test were as follows for the Vq group in the college sample:

	Before	After	Gain
VS program	53.75	56.38	2.63
SY program	57.33	58.17	0.84

The groups were not very equivalent prior to instruction (though not differing significantly, $t = 1.51$) but the gap was reduced by half through the greater transfer of the VS group. The direction of change was consistent with the hypothesis. Further, 6 out of 8 in the VS made some gain, in contrast with 3 out of 6 with the SY program, also tending to support the hypothesis.

The fifth hypothesis predicted greater transfer for the Qv group that studied class reasoning with the SY program version. The data for the Qv college students with respect to total scores in conditional reasoning were as follows:

	Before	After	Gain
SY program	62.38	63.63	1.25
VS program	58.29	58.00	-0.29

The group using the SY program did not gain much, but the VS group showed a slight decline. The direction of the gains is as predicted. Among those taught by the SY program 5 out of 8 gained on the transfer test, compared with 3 out of 7 in the VS group, again in the direction of the hypothesis.

The final hypothesis predicted greater transfer for groups assigned a program version matching their discrepancy type. Total scores for matched groups were as follows:

	Before	After	Gain
Matched	58.06	60.00	1.22
Mismatched	57.85	58.08	0.64

The direction of the difference in gains accords with the hypothesis, but statistical significance is lacking. Eleven out of 16 matched college students showed positive transfer, whereas only 6 out of 13 of those who were mismatched did. On the other hand, the analysis of covariance revealed no significant interaction effect.

Conclusions

1. The Cornell Class and Conditional Reasoning Tests can be used with college populations without a serious "topping" effect, yielding scores distributed approximately normally.

2. The Cornell Class Reasoning Test correlates about .75 with the Cornell Conditional Reasoning Test on a college sample.

3. No significant sex difference was found among college students in class or conditional reasoning.

4. No pattern of growth in deductive reasoning ability during the undergraduate college years was found.

5. Students reporting that they had taken or were taking one or more college courses in logic did not score significantly higher on either test, than students who had never studied logic.

6. Nevertheless, a brief 200-frame programmed instruction unit on class reasoning was able to produce significant improvement in class reasoning test scores on the part of college students.

7. Scores of college students on the Cornell Class Reasoning Test tend to be higher ($t = 7.00$; $p < .001$) than their scores on the Cornell Conditional Reasoning Test.

8. The mean college scores obtained (60.82 on class reasoning, 53.89 on conditional reasoning) were predictably higher than those reported by Ennis and Paulus for senior high school students.

9. If students do not systematically improve in deductive reasoning in college, and college students' scores are markedly higher than those of high school students, then apparently reasoning ability reaches its full development in high school and students with greater reasoning ability are selected for college.

10. Reasoning ability of college students is only moderately correlated (.31 -.36) with their academic achievement (GPA).

11. The discrepancy between a student's verbal and quantitative aptitude scores appears to be a variable related to his response to instructional materials.

12. Although differences were for the most part not statistically significant, twenty out of 27 indicators were in a direction supportive of the hypotheses that adaptation of instruction to direction of aptitude discrepancy facilitates learning of class reasoning principles and their transfer to conditional reasoning, only three indicators of the 27 being inconsistent with these hypotheses.

13. In several respects the procedures and findings of this study are similar to those reported by Baker (1968), who also used symbolic and semantic versions of a programmed instruction unit to teach logic to students classified on the basis of a battery of Guilford's ability tests. He found that success with both program versions was best predicted by scores on ability tests using semantic content and that aptitude-treatment interactions were not significant. Although his results were, like those of the present study, somewhat ambiguous, Baker found no evidence that the stimulus content dimension of programs is not critical from an instructional point of view.

Recommendations

1. Researchers concerned with deductive reasoning at the college level may find Ennis' Cornell tests, used in this study, to be suitable instruments for measuring this variable.

2. Further norming of the Cornell tests should be undertaken at a variety of institutions of higher learning to confirm the norms determined in this study and to ascertain what differences, if any, in reasoning ability exist among student bodies of various types of institutions.

3. Further refinement of the instructional programs on class reasoning developed in this study should be undertaken to provide an effective means of improving college and high school students' understanding of principles of logic. Parallel programs on conditional reasoning should be developed. Program revision and development should be in the direction of reducing the emphasis on principles which most mature students have been shown to understand and increasing the emphasis on the principles which are least understood.

4. Continued experimentation with aptitude-treatment interaction is to be encouraged. Although significant interaction was not demonstrated in this study, the direction-

ality of the findings was so consistent that the existence of an interaction effect seems at least to be possible. It is likely that the adaptation of instruction must be directed toward even more specific characteristics than the aptitude discrepancy used in this study. Further study is needed to determine what were the characteristics of those students who profited most from each of the program versions.

5. The personalogical variable of aptitude discrepancy appears to merit further study. The two extremes of this continuum appear to respond in noticeably different ways, even though their average I.Q's are similar.

6. Educators who are anxious to provide individuals with instructional materials consonant with their strengths should keep in mind that there are circumstances when an individual's needs are best served through materials which help him to overcome a weakness.

7. A study similar to that reported here using a program on conditional reasoning is needed to determine whether training in conditional reasoning transfers better to class reasoning than training in class reasoning does to conditional reasoning.

8. By developing a program parallel to the VS version used in the present study, using verbal examples from the humanities rather than the sciences it would be possible to determine whether an interaction effect exists with respect to the interest value of the content, in contrast with its form. The two verbal versions might be administered randomly to students majoring in the humanities and others majoring in the sciences.

9. In their efforts to adapt instruction to individual differences and to study various aptitude-treatment interactions, educational practitioners and researchers might well keep in mind the warning of Cronbach and Snow (1969, p. 194) that "... the Thurstonian approach [to aptitudes] which proved extremely useful in the field of vocational assignment (where persons are to do well at different tasks) has proved almost entirely abortive in guiding educational assignment, where we wish to bring persons by different means to master the same tasks."

VII

References

Andreas, Burton G. et. al. Research into Process Curricula. Eastern Regional Institute for Education, Syracuse, N.Y., March, 1970.

Baker, Robert L. An Application of Guilford's Structure of Intellect to Programmed Instruction. Office of Education, Bureau of Research, Project No. BR-5-0426, Arizona State University, February, 1968.

Berman, Louise. New Priorities in the Curriculum. Charles Merrill, 1968.

Bracht, Glenn H. "Experimental Factors Related to Aptitude-Treatment Interactions", Review of Educational Research. 40 (December, 1970), pp. 627-645.

____________. "The Relationship of Treatment Tasks, Personological Variables, and Dependent Variables to Aptitude-Treatment Interactions". Unpublished doctoral dissertation, University of Colorado, 1969.

Cronbach, Lee J. "How Can Instruction be Adapted to Individual Differences?", in R.M. Gagné (Ed.) Learning and Individual Differences. Merrill Books, 1967.

Cronbach, Lee J. and R.E. Snow. Final Report: Individual Differences in Learning Ability as a Function of Instructional Variables. Stanford University, March, 1969.

Dickhoff, James and Patricia James. Symbolic Logic and Language. McGraw-Hill, 1965.

Educational Policies Commission. The Central Purpose of American Education. National Education Association, 1961.

Ennis, Robert H. "A Concept of Critical Thinking", Harvard Educational Review. 32 (1962) pp. 81 - 111.

____________. Logic in Teaching. Prentice-Hall, 1969.

Ennis, Robert H. and Dieter Paulus. Critical Thinking Readiness in Grades 1-12. (Phase I: Deductive Thinking in Adolescence) Cooperative Research Project No. OE 1680, Cornell University, 1965.

Ennis, Robert H. et. al. Conditional Logic and Children. Cornell University, 1969.

Parker, J. Cecil and Louis J. Rubin. Process as Content: Curriculum Design and the Application of Knowledge. Rand McNally, 1966.

Rubin, Louis (ed.) Life Skills in School and Society. 1969 Yearbook, Association for Supervision and Curriculum Development, National Education Association, 1969.

Schagrin, Morton L. The Language of Logic. Random House, 1968.

Sharlow, John F. "Retroactive and Proactive Effects of Learning Certain Aspects of Symbolic Logic on 9th and 10th Grade Student Understanding of Mathematical Topics". Unpublished doctoral dissertation, State University of New York at Albany, 1971.

Smith, B. Othanel, Milton Meux, et. al. A Study of the Logic of Teaching. Cooperative Research Project, No. 258 (7257), University of Illinois, n.d.

Appendix A

Example of Frames from Program Version, SY

1.38 When the quantity term is "Some" instead of "No" or "All", the proposition is a particular instead of a universal. Because the affirmative statement, "Some A's are B's" does not assert anything about all A's, it is particular rather than ______________.

1.38 universal

1.39 Whereas the statement "No A's are C's" is universal and negative, the statement "Some A's are B's" is __________ and ________________.

1.39 particular
affirmative

Appendix A

Examples of Frames from Program Version, VS

8.3 The previous frame illustrated the contrapositive form which is true if the original statement is true. By interchanging the subject term and predicate term and also negating each of them, one can transform a statement into its ____________.

8.3 contrapositive (Don't confuse it with the converse where you merely interchange subject term and predicate term without at the same time negating them.

8.4 The contrapositive of "All electrons are negative particles" is all non-____________ ____________ are ______________-electrons.

8.4 All non-negative particles are non-electrons. (Negate both the interchanged subject and predicate terms.)

Appendix B

Sample Practice Item from Cornell Class Reasoning Test

Here is a reminder of the meaning of the possible answers:

A. YES It must be true.
B. NO It can't be true.
C. MAYBE It may be true or it may not be true. You weren't told enough to be certain whether it is "YES" or "NO".

Circle the answer to this next sample. Be careful:

3. Suppose you know that

Jane is standing near Betsy.

Then would this be true?

Betsy is standing near Jane.

3. A. YES
B. NO
C. MAYBE

Appendix C

Basic Data on College Sample

Subject	Sex	College Class	Logic Course	RSE Score V	RSE Score Q	Class Reasoning Pre	Class Reasoning Post	Symbolic Items Pre	Symbolic Items Post	Conditional Reasoning Pre	Conditional Reasoning Post
VV 1	F	72	No	83	26	61	62	8	9	47	49
2	F	72	Yes	86	29	64	67	10	10	61	64
3	F	71	No	90	28	64	68	8	10	59	59
4	F	72	No	83	17	54	49	5	8	51	53
5	F	71	Yes	83	14	58	62	9	8	57	56
6	F	71	No	83	33	59	61	8	8	50	60
7	F	71	Yes	90	23	67	69	11	11	56	57*
8	F	71	No	83	33	57	59*	8	8	49	53*
VS 1	F	72	No	84	29	68	68	11	11	60	61
2	F	72	Yes	84	23	58	54	8	7	54	53
3	F	72	No	85	28	67	67	8	11	56	56
4	F	71	Yes	83	14	61	67	8	12	53	58
5	F	71	Yes	88	24	57	61	5	8	56	57
6	F	72	No	88	27	63	68	8	12	65	64
7	F	72	Yes	88	18	52	62	11	11	**	**
QV 1	M	71	No	54	45	64	68	10	12	63	61
2	M	72	No	42	43	60	58	10	8	55	53*
3	M	70	No	52	46	65	60	9	7	53	54
4	M	71	No	53	45	62	66	8	11	61	59
5	F	70	No	58	49	68	68	11	12	59	61
6	M	71	Yes	52	44	65	63	11	9	53	51
7	M	71	No	59	50	68	72	11	12	64	47
QS 1	M	71	No	50	46	61	67	11	11	64	58
2	M	71	No	60	43	67	69	11	12	64	62
3	M	72	Yes	61	46	63	61	11	11	66	59
4	F	72	No	53	43	56	63	9	10	57	63
5	M	70	Yes	56	49	66	71	8	12	71	72*
6	M	71	No	54	48	68	71	9	12	62	65
7	M	72	No	61	44	61	66	7	10	55	65
8	F	70	No	53	45	64	70*	12	12	60	65*

* Eliminated from covariance analysis to equalize cells
** Did not take conditional reasoning test

Appendix C

Basic Data on College Sample (cont'd)

Score by Class Reasoning Principle

Subject		1[a]		2		3[b]		4		5		6		7		8	
		B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A
VV	1	6	6	5	6	10	12	5	6	4	6	6	6	5	6	4	6
	2	5	6	6	6	12	11	6	6	6	6	5	5	6	6	5	6
	3	6	6	6	5	12	11	6	6	6	6	4	5	5	4	4	4
	4	6	5	6	6	6	10	6	6	6	6	5	5	6	4	5	6
	5	6	6	5	6	11	12	6	6	6	6	5	5	4	6	6	6
	6	5	6	6	6	10	12	6	6	5	5	5	5	6	6	6	6
	7	6	6	6	6	9	10	5	6	6	6	6	4	5	6	5	6
	8	6	6	6	6	12	12	6	6	6	6	5	5	6	6	5	6
VS	1	6	6	4	6	12	12	6	6	6	6	6	6	6	5	4	5
	2	6	6	6	6	9	8	4	6	6	5	4	5	6	5	4	4
	3	6	6	6	6	10	9	5	6	6	6	5	5	5	6	5	3
	4	6	6	6	5	11	12	5	6	4	6	5	6	6	6	3	5
	5	6	6	6	6	11	12	6	6	6	6	5	6	6	6	6	6
	6	6	6	6	5	10	10	4	6	6	6	5	5	6	6	5	6
	7	6	6	6	6	11	12	6	6	6	6	5	6	6	6	6	6
QV	1	6	6	6	5	12	11	5	6	6	6	6	6	6	6	6	6
	2	6	6	5	6	9	7	4	6	6	6	5	1	4	5	5	3
	3	6	6	6	6	10	12	6	6	6	6	5	6	6	6	5	5
	4	6	6	6	6	11	12	6	6	6	6	3	3	5	5	4	6
	5	6	6	5	6	10	12	6	4	5	5	5	6	4	6	5	5
	6	6	6	6	6	10	11	5	6	6	6	4	5	5	5	4	6
	7	6	6	6	4	8	11	6	6	4	6	4	5	5	6	2	4
QS	1	6	6	6	6	7	9	6	6	5	5	3	3	5	5	6	6
	2	6	6	4	6	12	11	4	6	6	6	6	5	6	6	4	5
	3	6	5	6	6	11	12	6	6	6	6	5	5	6	6	4	5
	4	6	6	3	5	6	7	6	5	6	4	3	3	5	4	4	3
	5	6	6	5	6	8	11	4	6	6	5	4	2	6	6	5	5
	6	6	6	5	6	11	11	4	6	5	5	5	3	6	6	3	3
	7	6	6	6	6	11	12	4	6	6	6	6	5	6	6	6	5
	8	6	6	6	6	9	7	6	6	6	6	4	2	5	5	4	6

[a] Total possible for all principles except # 3 is 6.
[b] Total possible for principle 3 is 12.

B - Pre-test
A - Post-test

Appendix C

Basic Data on Ninth Grade Sample

Subject		Sex	CTMM Score L (V)	CTMM Score NL (Q)	Class Reasoning Pre	Class Reasoning Post	Symbolic Items Pre	Symbolic Items Post
VV	1	F	122	95	49	52	4	6
	2	M	123	110	49	58	7	7
	3	F	116	93	47	39	7	7
	4	M	115	94	50	54	6	6
	5	M	122	108	42	46	8	8
	6	M	120	105	34	48	4	7
	7	F	123	108	41	49	4	7
	1	M	132	103	61	67	8	11
	2	M	117	93	44	47	8	7
	3	F	129	111	59	56	5	7
	4	F	116	94	49	52	7	8
	5	F	119	98	43	49	4	8
	6	M	117	89	52	56	7	9
	7	F	118	103	52	48	7	6
	8	M	122	109	43	29	4	6
	1	M	110	129	63	64	7	8
	2	F	95	119	51	53	9	5
	3	F	108	131	54	57	8	9
	4	M	87	118	46	46	7	8
	5	F	101	119	54	59	8	9
	6	F	98	116	60	53	8	7
	7	F	104	120	57	60	8	8
	8	F	112	132	62	55	7	5
QS	1	F	104	118	55	51	7	5
	2	F	93	115	45	48	3	6
	3	F	93	117	58	60	8	7
	4	F	104	125	48	52	6	6
	5	F	97	117	59	53	7	6
	6	F	107	120	45	57	7	8
	7	F	100	119	43	43	5	7
	8	F	96	129	36	42	5	7

Appendix C

Basic Data on Ninth Grade Sample (cont'd)

Subject		Score by Class Reasoning Principle: 1[a] B	1[a] A	2 B	2 A	3[b] B	3[b] A	4 B	4 A	5 B	5 A	6 B	6 A	7 B	7 A	8 B	8 A
VV	1	6	6	6	6	6	5	5	6	5	6	1	2	3	3	3	4
	2	6	6	6	6	6	9	3	5	5	6	4	5	4	6	3	4
	3	6	4	6	6	2	2	5	4	6	3	1	0	2	1	5	4
	4	6	6	6	5	5	10	5	4	6	4	2	6	5	5	3	4
	5	6	6	3	6	4	5	3	3	6	5	4	4	3	1	4	3
	6	6	6	4	4	6	8	2	4	6	6	1	3	1	6	3	5
	7	6	6	4	6	8	6	4	5	5	6	2	1	4	3	4	3
VS	1	6	6	4	6	9	12	5	6	6	6	3	4	6	6	6	4
	2	6	6	6	6	3	0	5	6	4	6	0	0	1	1	5	5
	3	6	6	6	6	9	10	6	5	5	6	4	2	5	4	4	4
	4	6	6	5	6	6	3	5	6	5	5	1	2	4	4	6	6
	5	6	6	4	5	8	7	4	4	5	5	3	4	2	5	2	2
	6	6	6	4	6	7	8	5	6	6	6	5	4	5	6	4	5
	7	6	6	4	4	5	4	5	6	6	5	3	2	3	0	5	5
	8	6	1	4	3	5	2	4	3	5	4	3	1	3	2	5	4
QV	1	6	6	6	6	8	8	5	6	6	6	5	5	6	6	6	6
	2	6	6	5	5	6	8	4	5	6	5	3	3	2	2	5	6
	3	6	6	5	6	8	10	5	4	5	6	4	3	5	5	2	4
	4	5	6	5	5	7	6	5	6	5	5	3	0	4	3	3	5
	5	6	6	5	6	5	7	5	5	6	6	4	4	5	5	4	5
	6	6	6	6	6	11	9	5	4	6	6	2	2	5	4	4	4
	7	6	6	6	5	8	9	5	6	5	6	3	4	3	5	6	6
	8	6	6	6	6	10	10	5	5	6	6	3	2	5	3	5	2
QS	1	6	6	5	5	8	7	5	5	5	5	3	4	4	5	5	3
	2	6	6	4	5	6	5	5	5	4	4	2	3	2	3	4	5
	3	6	6	6	6	8	10	6	6	6	5	2	3	4	5	4	5
	4	6	6	5	5	5	8	4	4	6	5	3	4	3	5	5	5
	5	6	6	6	6	9	10	4	4	6	5	3	2	5	4	5	4
	6	5	6	6	5	8	10	5	5	4	5	2	4	2	2	2	3
	7	4	6	1	6	9	4	3	5	5	3	2	1	4	3	5	4
	8	5	6	4	5	2	2	4	5	6	5	1	0	1	2	2	4

[a] Total possible score for principles except # 3 is 6.
[b] Total possible for principle 3 is 12.

B - Pre-test
A - Post-test