DOCUMENT RESUME

ED 059 209 TE 002 790

AUTHOR Klein, Marvin L.; Grover, Burton L.
TITLE An Assessment of the Effectiveness of Symbolic Logic in the Teaching of Composition. Final Report.
INSTITUTION Cooperative Educational Service Agency 10, Plymouth, Wis.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-9-E-058
PUB DATE Oct 70
GRANT OEG-5-9-595058-0039(010)
NOTE 65p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Composition (Literary); Educational Research; English Instruction; Essays; *Logical Thinking; *Persuasive Discourse; Rhetoric; Secondary Education; Sentence Structure; Student Evaluation; *Symbolic Language; Teacher Attitudes; *Teaching Techniques; Writing Skills

ABSTRACT

This research project was carried out to test the hypothesis that symbolic logic instruction, if taught in a secondary school field situation in a way which included specific applications to argumentative composition, would effect greater improvement in composition and logical sentence analysis than if the students received little or no instruction in composition in an English program. Eighteen English teachers, in grades 9 to 12, from seven schools systems, taught symbolic logic to some of their classes. Seventeen replications of the study and three replications of a control class were obtained. The study produced 1264 pupil essays for scoring and analysis. Results of the study include: (1) logic instruction had a statistically significant effect on scores on the sentence logic test; (2) logic instruction produced no noticeable improvement on pupils' essay writing; and (3) twelve teachers were generally positive in regard to their plans to teach logic again. Five appendixes are included: A. Logic Test; B. Directions to Students for Essay Writing; C. Guidelines for Evaluation of Essays; D. Subgroup Mean Scores on Sentence Discrimination Test; E. Subgroup Mean Scores for Essays. The results of the data analysis are given in 16 tables. (DB)

FINAL REPORT
Project No. 9-E-058
Grant No. OEG-5-9-595058-0039 (010)

"AN ASSESSMENT OF THE EFFECTIVENESS OF SYMBOLIC LOGIC IN THE TEACHING OF COMPOSITION"

Marvin L. Klein
Burton L. Grover

Cooperative Educational Service Agency No. 10
111 East Mill Street
Plymouth, Wisconsin 53703

October 1970

HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report

Project No. 9-E-058
Grant No. OEG-5-9-595058-0039 (010)

"AN ASSESSMENT OF THE EFFECTIVENESS OF SYMBOLIC LOGIC IN THE TEACHING OF COMPOSITION"

Marvin L. Klein

Burton L. Grover

Cooperative Educational Service Agency No. 10
111 East Mill Street
Plymouth, Wisconsin

October 30, 1970

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 judgement 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

THE NATURE OF THE STUDY

Few people have ever contended that written composition instruction was an easy task. Indeed, it often appears that no other element of the curriculum presents as much frustration to teacher and learner at the same time. The act of writing brings to play constantly changing patterns of relationships between thought and knowledge, intuition and faith, personality and affective context.

When teachers do attempt to cope with the daily realities of composition instruction, they are often forced to turn to generalized vacuity implicit in terms such as 'unity', 'clarity', 'style', etc. Whether these concepts are consistently identifiable as lower order abstractions capable of citation and correction or improvement with application as cognitive and/or affective concepts is not a major point of contention. Our teaching is largely based upon the assumption that they are; otherwise our hopes for writing improvement would be even more dismal than at present.

Two more important questions need to be considered. First, is the operation and/or occurrence of lower level abstractions of a kind and order that they are generalizable throughout a broad domain of compositional tasks. Or, are they instead instances of stylistics in contexts so individual, in fact, that even to begin to cope with them requires far more linquistic sophistication than most of us will ever possess? Certainly, to the present time no one in any major study has apparently attempted to examine such element generalizability - perhaps because the potential appears so bleak.

Secondly, whether or not a wide ranging rhetorical generalizability exists, do these or other rhetorical concepts provide the basis for a "composition building or generating" instructional strategies which provide a means to teach writing other than the commonly employed techniques of analysing no evaluating previous writing. Current practices as exemplified in published materials, both textbook and other, continue to operate on a critical analysis premise, that is, through study of professional and/or peer writing models, students can ape successful writing techniques. They analyze the models—the teacher analyzes their writing.

The writing process determines the ultimate goal of the written product, whether or not one views this process as a mentalistic-intellective one or as a behavioral one. The matter of dealing with improved composition is then largely a matter of improving the process of composing.

The study at hand represented a probe into the domain raised by these two underlying questions. It attempted to identify, isolate and probe the potential of certain specific cognitive concepts as instrumental rhetorical tools in certain kinds of composition instruction. It was felt that there existed a major need for delimitation of goals and approaches--considerably more than had been attempted in most previous studies--in order to get at both particular individual aspects

of a rhetorical structure and the interrelationships of those aspects as they relate to total rhetorical unity. The study attempted to assess the learning effects of selected cognitive concepts which lend themselves to a composition building orientation rather than a composition analyzing-criticizing orientation.

Specifically, attention is directed to a rhetoric program embodying symbolic logic as a cognitive nucleus for instruction in argumentation and related persuasive exposition.

Symbolic logic has great flexibility in both mode of presentation and range of applicability. It is quite pertinent to typical linguistic expression and easily allows for presentation of a wide variety of the logical forms which English sentences exhibit. It enables students to note common words which determine these forms and their effect upon the semantics of sentences in which they occur. In addition, it provides a basis for explaining the relation of logical form to composition organization.

Born in the efforts of Whitehead, Russell and Carnap, symbolic logic represents a major stride toward development of an artificial symbol system whereby the basic character of syntax and semiotic meet; syntax and meaning, function and form, style and stylistics become as one.

In composition research done to date, most favorable reports appear to be resulting from modern linguistic theory. Specifically, generative-transform theory has appeard to be effective in some instances (note ch. II). Generative-transform theory has its built-in limitations, however—a sentence length syntax which, while extremely useful in sentence building, appears to be of little value as related to total essay effectiveness, other than whatever impact is carried by sentences as individual entities.

Symbolic logic, however, has a syntactic-semiotic range through various logical assertion types and their implications, e.g., equivalency form restatements-transposition to implication to application of De Morgan's Theorem. But it also has major implications for total essay impact by offering the simple valid argument forms in addition to extended complex argument forms for application as basic argumentation frameworks for expression.

Symbolic logic then has reference to the entire logical-semiotic-structural scope of argumentative exposition.

Perhaps most importantly it requires little in depth study to see the same underlying propositional concepts in the equivalency forms of symbolic logic, transposition and transferral descriptions of Jean Piaget and basic transformations in the generative-transform theories of Noam Chomsky.

The present study was based in part on elements of similarity between the two realms of propositional concepts.

The underlying question (Can the use of symbolic logic as taught by regular teachers who have had a short training session in the subject, improve the exposition writing performance of high school students in a manner discernible to persons not versed in symbolic logic) was transformed in the following hypotheses:

1. Instruction in symbolic logic to students on the high school level will improve the total organization of student's written compositions when evaluated according to guidelines based on five criteria (which are discernible to persons not trained in symbolic logic):
   a. Inclusion of clear assertions and valid processes in support of the assertions.
   b. Intellectual sophistication resulting from more mature and complex argument forms.
   c. Unified cognitive framework through valid logical relationships of work choice and assertion types.
   d. Rhetorical force through straight-forward style of logical argument which minimizes irrelevancies.
   e. Total effectiveness of writing by developing a reasoning impact through the above steps.
2. Instruction in symbolic logic to students on the high school level will improve students' sentence discrimination when presented with sentences which have four varying characteristics:
   a. Conciseness and clarity through work choice and control of syntactic expression.
   b. Structural variety of assertives, namely, greater selective use of conditionals, disjunctives, conjunctives, negatives and by-conditionals.
   c. Logical effectiveness through reduced ambiguity in the stance of the writer.
   d. Rhetorical force produced by intellectual tone.

In order to keep content within workable bounds, a decision was made not to include concepts of inductive reasoning, vindication models, and qualtification analysis as part of the instruction. The time factor precluded inclusion of the first two, and the complexity of quantification analysis was beyond the scope of the study.

However, instruction was designed to provide ability to handle the following basic concepts:

I. Deduction as process

II. Conceptual Sets

A. Truth-Falsity

B. Validity-Invalidity

C. Soundness-Unsoundness

III. Assertion types:

A. Atomic

B. Molecular

1. Conditional

a. Necessary

b. Sufficient

2. Bi-conditional

3. Disjunctive

4. Conjunctive

5. Negative

IV. Basic Argument forms:

A. Modus Ponens

B. Modus Tollens

C. Disjunctive Syllogism

D. Hypothetical Syllogism

V. Equivalency forms:

A. Transposition

B. Double Negation

C. Implication

D. Commutation

E. Conjunction Introduction

F. Exportation

G. De Morgan's Theorem

RELATED RESEARCH

Few would question the need for a strong viable rhetoric, one which is consistent, one which can involve the writer, one which both teacher and learner perceive as practical and useful. Yet, the research findings to date have offered little encouragement to teachers searching for patterns of instruction which offer much potential or bodies of content with any degree of intellectual respectability which can confidently be recommended for use:

> A review of the condition of language instruction reveals many opinions and some facts. The opinions cover the complete range from optimism to pessimism. The facts, to the extent that they are revealed by the status studies and error analysis, are almost uniformly discouraging. (Sherwin, 1969)

> Research in composition during the three year period [1963-1966] has fallen yet another step behind research in other fields. (West, 1967)

Studies, both major and minor, have documented evidence showing little relation between quantity of writing and quality of writing nor relation between intensity of evaluation and composition performance of students, (Burton and Arnold, 1963), (McColly and Remstad, 1963). Although the Burton and McColly studies operated within a high school setting, similar findings at the university level add further support to their observations. Dressel, Schmid and Kincaid examined writing performance of students at Michigan State College by comparing the writing of 99 students reporting the greatest number of writing assignments with the 99 reporting the fewest number of writing assignments. (Writing done in freshman communication was not included.)

Each student wrote a theme on an assigned topic at the beginning of the year and a final one at the end. Rated by two faculty members by means of a rating scale, the papers revealed that differences were insufficient to justify positive associations between the writing practice and improvements in writing, (Dressel, Schmid, Kincaid, 1952).

M. A. Christiansen reported similar findings in a doctoral study at Kansas University, (Christiansen, 1964).

These findings support other studies done over the past several years, (F. Heys, 1962), (Sutton and Allen, 1964).

This is not to imply, however, that new dimensions and perspectives have not been discovered. A Lokke and Wykoff study found that students' writing peaks early (after about twelve themes in one semester), (Lokke and Wykoff, 1948), this later supported by Scannell and Haugh, ((Scannell and Haugh, 1968).

The Scannell and Haugh study specifically set out to compare the effectiveness of objective composition tests as opposed to traditional

theme assignments in teaching composition. Using 306 tenth graders from two metropolitan schools, the experimenters set up control and experiment classes, the former with students writing a theme every other week, the latter administering objective multiple-choice tests over composition elements every week. Final writings revealed no significant difference in the two methods of instruction relative to writing improvement, although the authors warned of the limits of generalizing due to the short period of time (one semester) covered.

In similar studies dealing with specific kinds of student errors, Fellers found that college students continue to make the same kind of errors that they had made earlier as high school students in spite of continued instruction in composition, (Fellers, 1953). Bone found little correlation between errors in student themes and frequency of errors on a standardized test covering the same situation, leading to the possible inference that students avoid writing contexts which force them into situations requiring use of troublesome writing concepts, (Bone, 1964).

Some hopeful signs have been found in studies such as the one reported by Earl Buxton. He set up three randomly selected groups of college freshmen. One group did no writing, a second wrote a theme a week for a semester with papers evaluated, graded, and commented upon by the instructor, and the third group also wrote a theme a week with papers evaluated, graded, and commented upon, but in addition, thirty to forty minutes of class time was given over to discussion and revision. Although not exceeding the other groups in critical thinking, originality or organization, the revision group did show gains in rhetorical elements such as fluency and variety. Buxton used only two evaluators although there was a relatively high interrater reliability (.90), (Buxton, 1958).

McColly and Remstad in a study of the effectiveness of composition skills learning activities tested three hypotheses:

1) More writing alone results in better writing.

2) More functional "non-writing" composition learning activities result in better writing.

3) Tutoring with immediate feedback has significant benefit as a composition skills learning activity.

Involving slightly more than 300 students, grades 8 - 12, McColly condluded:

> The activity of writing in and of itself is fruitless... It can be inferred in part from the facts that writing per se is ineffective and functional instruction for about half the time spent in English is significantly more effective than one-fourth that amount. Until more is known, the best inference

> seems to be that to teach composition effectively, teachers should give a weekly writing task on which they base about 2 and ½ days of practical explanation, student practice, discussion, revising, rewriting, etc. (McColly and Remstad, 1963).

Attempts to relate writing improvement to the learning of particular cognitive skills and concepts have, for the most part, been unsuccessful also. Indeed, studies attempting to show relationships between a knowledge of formal grammar and an ability to write have been so discouraging over the years that it led a reviewer of research in the teaching of English to observe:

> Statistical and nonexperimental studies using correlation analysis by Hoyt, Rapeer, Boraas, Asker, Segal and Barr, Catherwood, Bradford, and Robinson failed to show a significant relationship between grammatical knowledge and writing ability. Except for Wykoff's study, the experimental studies by Briggs, Symonds, Crawford and Royer, Cutright, Ash, Benfer, Clark, Warner and Guiler, Milligan, Frogner, Krause, Smith, and Maize also failed to support the case for grammar. After a tally of procedural and other limitations, the research still overwhelmingly supports the contention that instruction in formal grammar is an ineffective and inefficient way to help students achieve proficiency in writing. (Sherwin, 1969).

Again, however, this is not to assert that some promising findings have not occurred. Especially encouraging have been studies based upon generative-transformational grammatical theory. Kellogg W. Hunt's monumental study of the development of syntactic structure at three grade levels (4, 8, and 12) provides insight into the kinds of syntactic sophistications normally accruing in the maturing writer, (Hunt, 1964). But more importantly, Hunt has introduced techniques of analysis such as T-Unit measures based upon work of Chomsky and Lees which have enabled further studies to more effectively measure syntactic fluency and sophistication.

The result has been some attempt to find specific effects of studies of transformational grammar on the writing of students. For example, the Bateman and Zidonis study set up experiment and control classes involving fifty students at Ohio State University in the ninth and tenth grades, (Bateman and Zidonis, 1966). The two year project was premised upon the idea that a study of special transformational materials would effect more "well-formed" sentences from the experimental group. According to the authors, "To be considered 'well-formed', a sentence had to be both intuitively acceptable to the analysts and derivable from the rules of the grammar."

Using a technique of syntactic complexity analysis based upon transformational theory, they concluded "A knowledge of generative grammar enables students to increase significantly the proportion of well-formed sentences they write." They also concluded, "A knowledge of generative grammar can enable students to reduce the occurrence of

errors in their writing."

In a study published in 1969, John C. Mellon investigated the relationship that exists between practice in combining separate kernel sentences into single statements and the ability to produce more structurally elaborated sentences. Concentration of analysis was upon what Mellon refers to as "syntactic fluency," that is, "average structural diversity in terms of the frequency and depth of nominal and relative embeddings." (Mellon. 1969).

Involving 247 seventh graders, Mellon observed that "systematic programs entailing the a-rhetorical intensive, and specially planned experiencing of mature sentences will increase the rate at which the sentence structure of the student's own writing becomes more highly elaborated (or differentiated) and thus more mature." He further contends that "it was the sentence-combining practice associated with the grammar study, not the grammar study itself, that influenced the syntactic fluency growth rate." It appears that here he might be on questionable ground since the delineation between those two aspects of the performance aren't as isolable as one might prefer.

Although Mellon is quick to point out that sentence-combining practice has nothing to do with the teaching of writing, many will undoubtedly equate increased syntactic fluency with improved writing. As an aspect of the experiment, Mellon, using six junior high school teachers as evaluators, attempted to measure writing performances especially in the areas of ideas, organization, style, sentence structure, and vocabulary.

Using a five-level scale based upon the techniques of rapid-reading developed by the College Entrance Examination Board, Mellon observed,

> Subsequent pair comparisons showed that the standing of the controls was also significantly higher than that of the experimentals, but that the experimental and placebo groups were indistinguishable. Strictly speaking, then the question raised in the overall quality comparison was answered ambiguously. The writing of the experimental group was inferior to that of the subjects who had studied conventional grammar, but indistinguishable from that of subjects who had studied no grammar but had received extra instruction in composition-curious results indeed. (Mellon, 1969).

This, in spite of the fact that there was an inter-rater reliability of .83 and, in Mellon's opinion, "The raters' judgements appeared to be both consistent and valid."

Conclusions and Implications

Conclusions

A) There is little reason for elation in the findings of most major research in composition to date. Major aspects of the traditional English curriculum have failed to produce expected results. Formal grammar instruction, major amounts of writing practice, increased intensity of writing evaluation have all failed to produce better writers. There is, however, some research support for motivation, discussion, explanation and revision as phases of a functional instructional program in composition.

B) Considerable difficulty appears to remain in the whole definitional range of "good writing." Is "good writing" made up of "well-formed" sentences or "syntactically fluent" sentences? Is the presence of either or both of these a sufficient or necessary condition of "good writing"? Can inter-rater reliability ever be such that concurrence is implicit in individual factor as well as total essay aspects?

There seems to be little evidence to support a generalized consensus of "good writing" without at least some reference to specific aspects of subject, audience and purpose.

C) Most favorable results to date appear to be in the area of increased sophistication of linquistic analysis. Specifically, techniques derived from generative transformational theory have provided significant evaluative insight. In addition, however, they have brought with them the built-in limitations of that same theory-- a sentence length syntax which, while extremely useful in sentence building and analysis, appears to be of little value as related to total essay effectiveness, other than whatever impact is carried by sentences as individual entities.

D) Many instructional strategies and techniques as revealed in research studies and various published materials continue to orient toward product and analysis rather than process and generation. In many respects, the product is an after-the-fact aspect of the composing act. Most favorable results instead appear to be coming from "composition building" instructional strategies.

E) Few studies attempt to isolate specific cognitive concepts as significant rhetorical variables in the writing process. Nor has there been any substantial attempt to relate any such variables to such aspects as mode of discourse, nature of thesis, writer purpose, audience, etc.

Implications

If there is hope for the potential value and usefulness of a cognitively oriented rhetoric, there appears to be a major need to: 1) discover the nature of the cognitive concepts of that rhetoric, both as individual aspects of a total rhetorical unity and as interrelating variables contributing to the total rhetorical unity 2) probe possible patterns of relationship existing between those congnitive concepts and the various affective phases and aspects of a rhetoric 3) develop and test models of a prototypic rhetoric with conceptual relationships delineated enough to provide handles for evaluation and analysis of concepts during composition instruction.

There is certainly reason to explore possible areas which might assist in shedding light on any of these three needs. for it is in this realm that current work and evaluative techniques have made the most progress, and, at least to the present time, offer the strongest sense of operational viability.

PROCEDURES

Description of the Problem and Context of the Field Situation

The basic problem of the study was this: Given a group of secondary school English teachers, trained in symbolic logic program over a two-week period and teaching the material to fairly typical English classes over a four-to-seven-week period, would special instruction in symbolic logic effect improvement in argumentative writing performance and in ability to analyze sentences logically? It was hypothesized that the symbolic logic instruction, if taught in a secondary school field situation in a way which included specific applications to argumentative composition, would effect greater improvement in composition and logical sentence analysis than if students had an English program that in most cases had little or no instruction in composition.

The field situation in which the hypotheses were tested provided a special context for the study. Regular secondary school English teachers who had volunteered for the project and for the in-service training taught the material to some of their regularly assigned classes. The material was used in grade levels 9 through 12 in various schools. Instruction of the material to the pupils varied from four to seven weeks, depending upon the judgement of the teacher and restrictions imposed by specific school situations. The average scholastic ability of the classes selected for the experiment were in most cases in the average range; there was some attempt to eliminate low average classes from the experiment. The experiment attempted to assess differences among grade levels and between the writing of themes in-class or as out-of-class assignments as well as the effectiveness of the logic program.

The research project is perhaps best characterized as a field experiment which attempted to evaluate the application of an innovative program for the teaching of composition. In a sense, it constituted an evaluation of an in-service program for teachers with the criterion two steps removed from the program. Teacher understanding of the material was not assessed directly--nor was subsequent teaching performance-- rather the training of teachers was evaluated by assessing subsequent effects on pupil writing. In order for the hypothesis to be supported, not only would the symbolic logic as adapted for this program have to have a potential effect on writing, but the in-service program and materials supplied to the teachers would also have to be such that the potential effect of symbolic logic on pupil performance would be realized. This nature of a field experiment provided a more stringent test of the hypotheses than in a more carefully monitored and controlled laboratory situation, but verification in such a field situation would supply greater confidence that the program would be applicable over a wider range of teachers and schools. What it lacked in precise control, it gained in external validity.

Preparation of Teachers

With the cooperation and support of seven school districts which paid stipends to teachers, seventeen secondary English teachers volunteered for the project. They attended a two-week training session in symbolic logic taught by Marvin Klein during August 1969, before the start of school. There were ten full day sessions, seven hours each, during this two-week period. With the exception of a few hours devoted to the design and procedures of the experiment, the full time was devoted to the study of symbolic logic (for the teaching of composition) as specifically adapted by Marvin Klein in Symbolic Logic and Its Application to Writing. Prior to attendance and selection for the workshop, each teacher agreed to teach the logic from a four- to seven-week period at different times to two of their regularly assigned classes and to withhold logic instruction from a third class. All but one teacher (who was not included in the final statistical analysis) found it possible to teach the logic to two classes. Thirteen teachers were able to include a third control calss in the experiment. For the remaining three teachers, a separate control class taught by another teacher had to be identified and used; provisions were made to analyze these three situations separately, should the initial descriptive analysis have warranted such special separate analysis.

Population

The study collected data from 1264 students who were assigned to the classes of the teachers in the study by regular school scheduling procedures. The classes of the teachers ranged from grades 9 to 12. Partial information suggested that class averages of scholastic aptitude or intelligence were in the average range or slightly above average. An attempt was made to eliminate the low average classes from this study. The students in the classes were all or almost all white and the school systems involved in the study were those of small cities, towns, and rural areas of east-central Wisconsin.

Classes were distributed over four grade levels and seven school districts. Seventeen teachers who were able to teach logic to two classes included four ninth grade teachers, five tenth, five eleventh, and three twelfth grade. All teachers had two classes on the same grade level to which they were able to teach the material. In three cases a third control class taught by another teacher had to be found in the same grade level and school system for those teachers in the study. Table I indicates the distribution of classes by grade level and school system. Because the design permitted a replication of the experiment for each teacher, no attempt was made to assess between-school differences in the analysis. Between-school differences were extraneous to the study, and such an analysis would unduly complicate and unbalance the design.

Measures

In order to measure pupils' ability to analyze the logical argument contained by various forms of sentences, an eighteen-item test was constructed. All items were multiple choice. Items asked students to identify implied conclusions in argumentative sentences, to identify conclusions within sentences, and to identify correct logical translations of argumentative sentences. This test is appended as Appendix A.

In order to assess pupils' ability to write clear and logical arguments, they were asked to write essays on a topic of their own choosing. Some possible topics were suggested to them. No more than one essay was collected from any one pupil in the study for the purpose of the study. The set of directions given to students for essay writing included as Appendix B.

Each essay was read by four raters. No rater had any formal background in symbolic logic. The experiment was designed to test the possibility that people who were themselves not versed in symbolic logic could detect superiority in writing occasioned by previous training by the writer in symbolic logic. A rating scale was devised to be used by each rater of each essay. The rating scale consisted of the following nine factors: (concerning sentence structure) clarity, structural variety, logical effectiveness and rhetorical force; (concerning organization) soundness, sophistication, unity, rhetorical force, total effectiveness of paper. Each of the nine factors was arbitrarily assigned a maximum point value ranging from five to twenty-five. The maximum point value that could be assigned to any essay was onehundred. Guidelines which explained the intended meaning of the factors as well as how to apply them to the rating of essays and which were used in the training of raters, are included as Appendix C.

In all, nine raters were employed by the study. Any one essay was read by four raters and not the total group of nine. In order to eliminate any systematic bias arising from differences among raters, each rater went through a subset of essays, designated as a "batch," at a time. Each batch consisted of one essay from each of 153 subgroups in the experiment. If differences did exist among the raters on the general overall rating of the essays, it would affect no experimental subgroup more than another save for a possibility of a very slight difference associated with incomplete batches resulting from those subgroups that included more students than others.

Raters were given two practice sessions in the rating of themes. Instructions from the principal investigator were given them about the meaning of the factors to be used in the rating scale as well as the procedures used in rating and sorting and passing around essays. A couple trial runs in rating essays were made to check inter-rater reliability before proceeding with the main body of essays. During the trials, inter-rater reliability was computed using an analysis of variance procedure. Hoped for inter-rater reliability of .85 for four raters (average of four raters' scores) was not achieved. Rating

of essays was begun after raters had attained reliability of .77 on a trial run (an average reliability coefficient between any two raters of .46 would yield a reliability of .77 for the average or total scores of the four raters combined).

Design

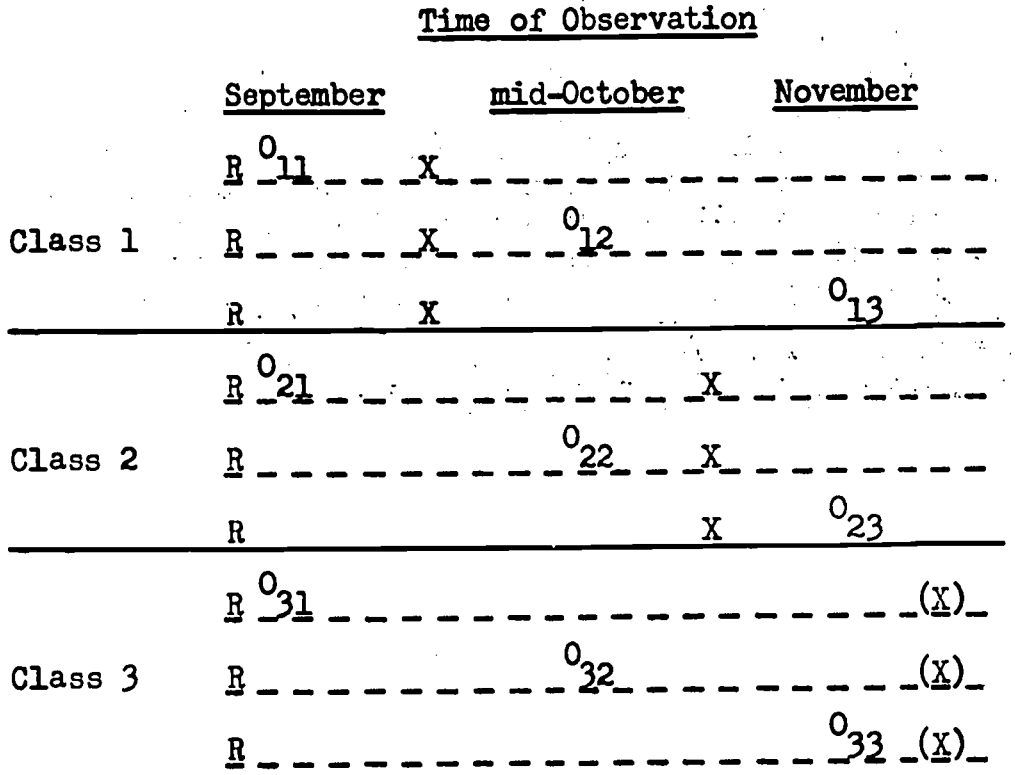
The design for testing the hypotheses had to be arranged within the context provided by the field situation, the nature of the instruction, and constraints provided by the ongoing operation of the public schools involved. This context was partly responsible for certain basic decisions about the design: (1) individual pupils were not considered as experimental units and intact classes would be kept intact for purposes of instruction, (2) there were no repeated measures on individuals because of a suspected reactive effect if too much criterial writing was asked of each pupil; there were, however, repeated measures on classrooms by testing different subgroups within the class at different times, (3) separate school systems were not to be considered as an accounted-for factor in the design even though some confounding of school system with grade level would result, (4) teachers were given the option of varying the length of instruction between four to seven weeks as well as making all day-to-day decisions about instruction, and (5) where a teacher could not provide a third control class on the same grade level, a control class taught by another teacher would be secured; in cases where other types of composition instruction could not be withheld from the control class, this would be noted; if warranted by preliminary inspection of the data, a separate analysis would be performed on the replicates which had these special control classes.

A basic design was worked out for the three classes of each experimental teacher in the study, and each teacher constituted a replication of the experiment. One class was randomly selected to receive logic instruction during the first 4-7 week period of the school year, one class randomly selected to receive logic instruction during the second 4-7 week period, and the third class to receive no logic instruction. When not receiving logic instruction, instruction in writing was in most cases withheld from other classes so that a basic comparison with instruction unrelated to writing could be made.

For testing purposes, each class was divided randomly into three groups. One group was tested at the start of the year, one immediately after logic instruction was completed in the first class to receive it, and one at the conclusion of the experiment which coincided with the conclusion of logic instruction for the second class to receive logic instruction. Although the pupils of each testing subgroup were different groups within a class at different times constitutes a repeated measure on a class; from this, gain scores of classes over the period of time that logic was taught could be compared with gain scores of classes when logic was not taught or had not been taught. Two gain scores associated with experimental instruction were compared with gain scores associated with "control" instruction. Three gain scores associated with control instruction were obtained, two from the third

control class for the two time periods of the experiment and one from the first time period of the class which received logic instruction during the second time period.

Using "O" for test or observation, "R" for randomization, "X" for instructional treatment in logic and "(X)" for instructional treatment outside the study, a solid line for division of intact classes, and a broken line for division of randomized subgroups, the design for one block may be illustrated as follows (To repeat, which of three selected classes would become Class 1, Class 2 or Class 3 was determined at random assignment):



According to this design, two classes of any of the teachers in the study received the same instruction, but because they received the instruction at different times and because the testing occured at intervals before and after instruction, it was possible to make internal comparisons about relative growth that tested the hypotheses of the study, specifically insofar as symbolic logic's effectiveness compares to instruction not directly related to composition.

The basic test of the effectiveness of the symbolic logic instructional package was by means of the following planned comparison summed over seventeen replications. In the comparison below, any "O" represents the average score of the subgroup summed over seventeen replications with the subscripts corresponding to the chart above

$$[(O_{12} - O_{11}) + (O_{23} - O_{22})] - 2/3\ [(O_{22} - O_{21}) + (O_{32} - O_{31}) + (O_{33} - O_{32})]$$

which simplifies to

$$O_{12} - O_{11} + O_{23} - 1\ 2/3\ O_{22} + 2/3\ (O_{21} + O_{31} - O_{33})$$

It had been intended that there would be twenty replications distributed equally over four grade levels. With this distribution it would have been possible to have a balanced design with respect to grade levels, which would ease the task of obtaining an error term for the basic comparison from an analysis of variance. However, only seventeen teachers were able to carry out the experiment, and distribution within a grade level ranged from five to three (Table I).

At a later stage in planning, another factor was added to the design. This factor was whether or not pupils were to write criterion essays during class time (given two class periods for organization and writing) or as an out-of-class assignment. Either procedure seemed to have potential advantages and disadvantages. Out-of-class writing suits the nature of the assignment and makes it easier and more natural to apply what had been learned about logic. However, Out-of-class writing seemed to lack the control of testing provided by in-class writing. Besides inter-pupil cooperation on the essays, out-of-class writing could allow a motivational factor affecting effort and time spent to run rampant—and two very possible affected dependent variables, cognitive writing ability and motivation to write, would be confounded—and yet interact with instructional variables in different ways. To avoid this dilemma, teachers were randomly divided into two groups on each grade level, one group to have pupils write out of class and one group to write during class time.

RESULTS

Measurement of Writing Performance; Inter-rater Reliability

The total study obtained 1264 essays from pupils. Four raters from a pool of nine raters rated each essay. Inter-rater reliability was assessed by a rater by essay analysis of variance procedure. With the set of raters varying from essay to essay, raters of the set for any essay were arbitrarily designated one, two, three, and four; and the analysis considered all who were designated with the same number as the same rater. This artifact did not affect the meaning of the obtained inter-reliability coefficients for the purposes of this study.

The analysis for inter-rater reliability is presented in Table II. The estimated average correlation between any two raters was r = .46. The estimated reliability of an essay score which was a total (or average) of four individual rater scores was r = .77, the same coefficient that was obtained from the final trial run during the training of the raters.

Measurement of Writing Performance; Factor Analysis of the Essay Scores

The total score for an essay (whether from a single rater or summed over four different raters) was a sum of nine subscores. The nine subscores from the total set of 1264 essays (each summed over four raters) were analyzed with a principal axis factor analysis program to see if in fact the subscores constituted different factors or if any groupings of them were actually independent of one another.

To put it bluntly, they weren't.

All subscores were intercorrelated to a fairly high degree with the intercorrelations ranging from .50 to .92. The factor analysis retained only one factor, and there was no subsequent varimax rotation. The loadings of all subscores on that one factor ranged from .82 for clarity to .96 for rhetorical force. The intercorrelations and factor loadings are included in Table III.

The factor analysis provided little or no empirical justification for the belief that the subscore labels represented different things, despite what the investigators or the raters may have had in mind during the study. In fact, any one of three of the variables (logical effectiveness, sophistication, and rhetorical force), at least as scored within the context of the task of deriving all subscores, could have been used in place of the total score with very little change in results. Whether or not the scoring would have been altered if no attention had been paid to any subscores is hard to say, but the use of the nine particular subscores in rating themes in this particular study turned out to be a measurement of one factor and one factor only. (On rational grounds, it could be argued that the factors are still different things even though they are highly intercorrelated; even if this were a more accurate interpretation, it nevertheless turned out that a measurement of one was also a measurement of another.)

T A B L E I

Distribution of Replications by School System, Grade Level, and Testing Conditions

CR - Complete replicate: one teacher and three classes
PR - Partial replicate: one teacher and two classes receiving logic instruction
CC - Control class for a partial replicate
ic - in-class writing test condition
oc - out-of-class writing test condition

Eastern Wisconsin School Systems

		Two Rivers	Sheboygan	Plymouth	Reedsville	Manitowoc	Elkhart Lake	New Holstein
9th Grade	ic	CR		CR				
	oc					CR		CR
10th Grade	ic			CR	CR	CR		
	oc					CR CC	PR	
11th Grade	ic	CR	CR PR CC					
	oc		CR PR			CC		
12th Grade	ic					CR		
	oc	CR	CR					

T A B L E II

INTER-RATER RELIABILITY OF ESSAY SCORING BASED ON RATER-BY-PUPIL ANALYSIS OF VARIANCE

SOURCE	d.f.	SS	MS
Rater	3	1504.	501.3
Pupil	1263	1,336,336	1058.6
R x P	3789	923,328	243.69
Total	5055		

$$r = \frac{\text{MS pupil} - \text{MS p x R}}{\text{MS pupil}} = \frac{814.37}{1058.06} = .77$$

(The four raters for any one essay varied from essay to essay among a set of nine raters. The raters of any essay were designated with a number from one to four. This artificial coding of raters does not affect interpretation of resultant inter-rater reliability.)

TABLE III

INTER-CORRELATIONS OF SUBSCORES OF ESSAY RATINGS (USING SCORES FROM 1264 ESSAYS, EACH SCORE BEING A TOTAL FROM FOUR DIFFERENT RATERS)

Sentence Structure	1	2	3	4	5	6	7	8	9	Loadings of Subscores on Lone Factor From Factor Analysis
1. Clarity	1.00									.825
2. Structural Variety	.804	1.00								.896
3. Logical Effectiveness	.835	.864	1.00							.938
4. Rhetorical Force	.790	.834	.867	1.00						.917
Organization										
5. Soundness	.504	.621	.729	.664	1.00					.837
6. Sophistication	.795	.863	.910	.886	.738	1.00				.950
7. Unity	.605	.697	.701	.700	.795	.748	1.00			.852
8. Rhetorical Force	.771	.846	.883	.909	.778	.922	.809	1.00		.963
9. Total Effectiveness	.511	.651	.722	.687	.914	.745	.809	.805	1.00	.849

T A B L E IV

MEANS AND STANDARD DEVIATIONS OF SUBSCORES
OF ALL ESSAY RATINGS (N = 1264)

Criterion Rated	Highest Allowable Rating	Mean	Standard Deviation
Sentence Structure			
1. Clarity	10	7.37	1.95
2. Structural Variety	10	6.56	1.82
3. Logical Effectiveness	10	6.10	2.03
4. Rhetorical Force	5	2.85	.99
Organization			
5. Soundness	25	15.34	4.19
6. Sophistication	10	5.57	2.00
7. Unity	10	6.35	1.79
8. Rhetorical Force	10	5.48	1.87
9. Total Effectiveness	10	5.87	1.69

Sentence Discrimination Results

Pupils' understanding of the material insofar as sentence structure was concerned was assessed in a rather direct way by an 18-item multiple choice test. The data analysis provided marginal support for rejection of the null hypothesis that instruction had no positive effect of pupils ability to discriminate sentences according to logical structure.

For fourteen of the seventeen replications, the basic contrast was in the predicted direction; in other words, more gains were associated with the times that logic was taught than with the times when it was not taught. This was considered sufficient grounds for rejecting the null hypothesis; the probability that a difference could have occurred in the predicted direction 14 out of 17 times by chance is 0.6%.

Strangely enough, the ordinarily more sophisticated, powerful parametric test, using an error term obtained from the analysis of variance, provided somewhat less basis for rejection of the null hypothesis, probably because the average of the three negative results were larger than the average of the 14 positive ones. Nevertheless, this test also was significant at less than the 5% level.

The means of the subgroup means from each of the 17 replicates is presented below in Table V.

TABLE V

SUBGROUP MEAN SCORES ON OBJECTIVE TEST AVERAGED OVER ALL 17 REPLICATES

	Testing Time		
	1st	2nd	3rd
Logic First Group	9.04	10.03	10.15
Logic Second Group	9.11	9.54	10.66
No Logic Group	8.59	8.84	8.90

The average gain when logic was taught in the logic first group was .986 and in the logic second group 1.010. Averaged the two groups together, the gain was .998. This compared with an average gain of .378 for the three time periods used as a control. Using an error term obtained from an analysis of variance, the obtained t value was +1.85,

significant for a 1-tailed test at the .05 but not at the .025 level (Table VI). (The error term obtained from the analysis of variance was a pooled term of interaction of instructional sequence and testing time with replications; placing instructional sequence and testing time into rows and columns as in Table V indicates that a positive contrast would reflect itself not only as an interaction but also as row and column effects, hence a pooling of replication interactions seemed appropriate).

T A B L E VI

BASIC CONTRAST USING CELL MEANS OF TEST SCORES

Experimental Time Period Gains:	.986 + 1.010	= 1.996
Control Time Period Gains:	2/3[.595 + .380 + .160]	= .756
Sum of Contrast Coefficients Squared	7.111	
Pooled Mean Square Error:	1.07	
Number of Replications	17	
Number of Scores per Cell Per Replication:	1	

$$t = \frac{(1.996 - .756)17}{\sqrt{(1.07)(7.11)(17)}} = \frac{21.080}{\sqrt{129.35}} = \frac{21.080}{11.37} = +1.85$$

Appendix D presents means for each subgroup according to each replication, and the resultant contrast figure for each replication are presented in Table VII. The three negative or "wrong" contrasts all occurred at the 10th grade level in three different school systems. The most positive contrasts, on the average, occurred in grade nine. This pattern of contrasts by grade level is not easily explained by anything other than chance.

T A B L E VII

CONTRASTS OF TEST SCORES BY GRADE LEVEL

Grade	Contrasts For Each Replication				
9	17.41	7.11	13.32	2.17	
10	-21.22	-7.27	1.84	1.52	-8.95
11	4.19	9.81	13.90	4.45	.23
12	4.49	.29	13.92		

An analysis of variance of the data was performed to obtain an error term for the basic contrast and to gain supplementary information about grade level and other effects (such as in-class vs out-of-class writing of the essays). A summary of the analysis is presented in Table VIII. To have a balanced design for the analysis, one replication was dropped randomly from the 10th and 11th grade groupings and figures estimated for a dummy replication on the 12th grade level. Grade level differences in the analysis did not reach usually acceptable levels of significance in this analysis, but the test was very conservative for this effect in the use of the between-replications (within grade level-test condition cells) factor on an error term. A description of average scores per grade level is perhaps more revealing (if one keeps in mind that the replications were selected and not sampled from a larger population). The ninth grade average of subgroup means was 8.71, tenth grade was 9.07, eleventh was 9.98, and twelfth was 9.68.

TABLE VIII

ANALYSIS OF VARIANCE OF OBJECTIVE TEST SCORE DATA

Source	d.f.	SS	MS	F
Grade Level (G)	3	36.3	12.1	2.52
Essay Test Condition (W)	1	3.8	3.8	.80
G x W	3	16.0	5.3	1.11
(a) Replications W/GW	8	38.4	4.8	
Class W/R or Instructional Sequence (C)	2	30.7	15.3	10.37
G x C	6	12.1	2.0	1.36
W x C	2	2.3	1.2	.79
G x W x C	6	18.4	3.1	2.08
(b) R x C	16	23.7	1.5	
Testing Time (T)	2	23.7	11.9	13.24
G x T	6	19.7	3.3	3.66
W x T	2	2.7	1.3	1.50
G x W x T	6	4.7	.8	.87
(b) R x T	16	14.3	.9	
C x T	4	9.6	2.4	2.54
G x C x T	12	19.8	1.6	1.74
W x C x T	4	.9	.2	.23
G x W x C x T	12	11.3	.9	1.00
(b) R x C x T	32	30.3	.9	

(a) Assuming replication to be a random factor, MS_{Rep} used as error term for effects listed above it in table.

(b) Error term for basic contrast was a pooling of these three sources of variance.

Other than that which is revealed by the contrasts, the clearest description of the pattern of subgroup means is probably that presented by Table IX in which a frequency distribution of means is plotted and average means by grade level test times, and class or instructional sequence are presented for comparison on the same scale.

T A B L E IX

GRADE LEVEL, TESTING TIME, AND CLASS (INSTRUCTIONAL SEQUENCE) AVERAGES COMPARED TO TOTAL DISTRIBUTION OF SUBGROUP MEANS

Grade Level Averages

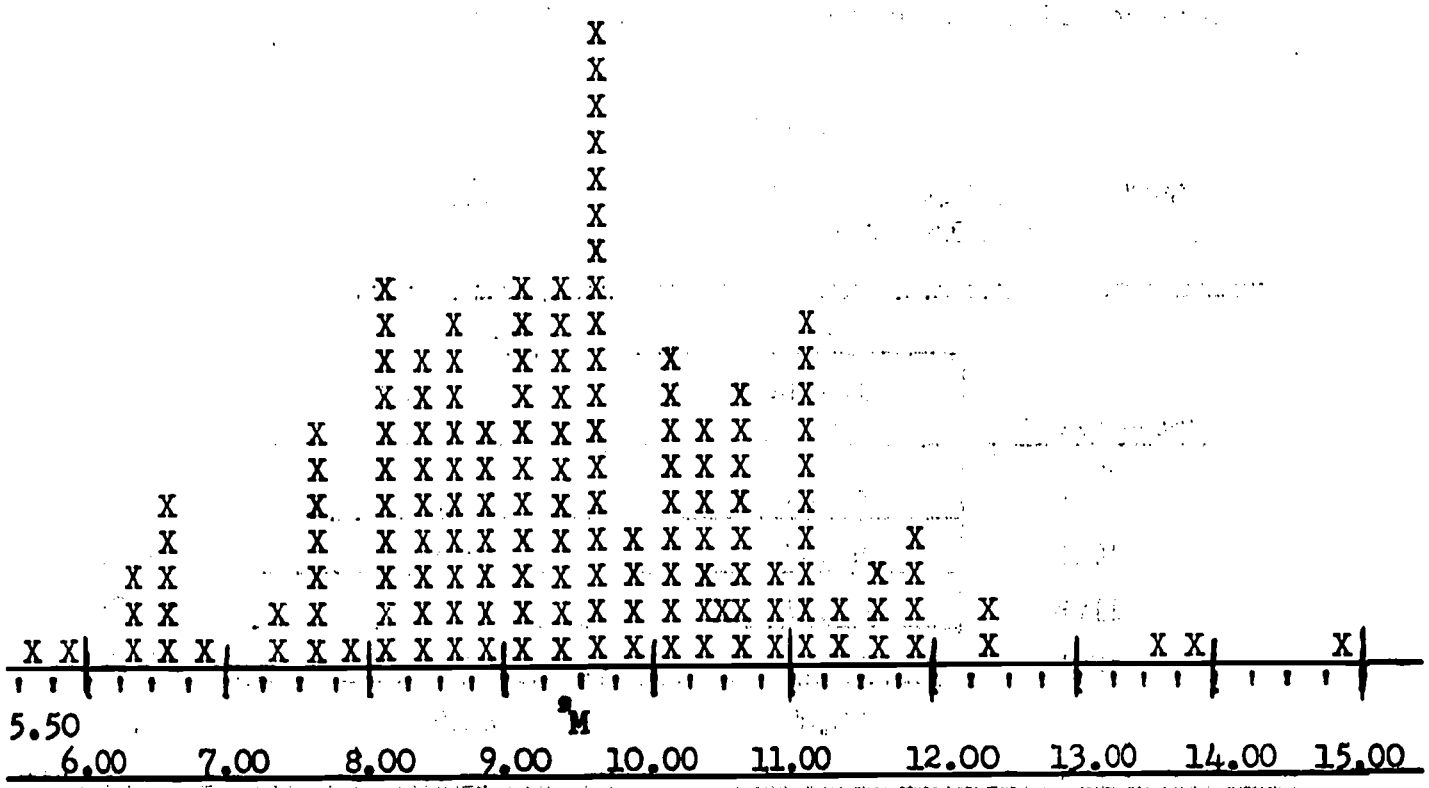
12th	9.68
11th	9.98
10th	9.07
9th	8.71

Time of Testing Averages

1st	8.91
2nd	9.47
3rd	9.90

Class (Instructional Sequence) Averages

Control	8.78
L - 1st	9.75
L - 2nd	9.75



Distribution of all Sub-group Means (M = 153)

Grand Mean = 9.42

For all groups the average was 9.42 or slightly more than 50% correct. The standard deviation was 1.49. The average gain on the test when logic was taught was .998 or roughly two-thirds of a standard deviation. The average control gain was .378 or roughly one quarter of a standard deviation. The net gain associated with logic instruction on the test was therefore approximately 5/12ths or .42 standard deviation units.

Essay Writing Results

The data from the analysis of essay tests supported the null hypothesis of no effect on essay tests associated with instruction in logic. There were more replicates (11) in which control gains exceeded gains associated with logic instruction (each resulting in a negative figure for the basic contrast) than replicates with positive results (6). This was sufficient evidence for acceptance of the null hypothesis when tested against the one-tailed alternate hypothesis of positive gains associated with instruction.

The average score assigned to each essay (or more precisely, the average of the subgroup averages of essay scores) was 61.48. The subgroup means ranged from 41.2 to 85.7, and the standard deviation of these means was 9.17. The standard deviation of individual pupil essay scores (averaged over four raters each) was 32.5.

The grade level averages (where means were obtained from an analysis which had dropped two replicates randomly to achieve a balanced design) and test condition averages are presented in Table X.

TABLE X

ESSAY SCORE MEANS OF SUBGROUPS FOR EACH GRADE LEVEL AND TEST CONDITION FOR 15 ACTUAL AND ONE ESTIMATED REPLICATE IN STUDY

	Test Condition		
Grade Level	In-Class Writing	Out-of-Class Writing	
9th	57.86	56.72	57.29
10th	59.37	56.31	57.84
11th	61.72	69.74	65.73
12th	64.53	68.98	66.75
	60.87	62.94	61.90

The grade level averages went up for each succeeding grade from 9th to 12th. There was a hint of an interaction of test condition with grade level with out-of-class writing associated with higher scores in 11th and 12th grade and with lower scores in 9th and 10th. None of these differences, however, were statistically significant at usual levels according to an analysis of variance which used between-replicate variance as an error term for these factors (Table XIV). Because the teachers qua replicates had not been randomly sampled from a larger population, the use of this error term--or any test of significance, in fact--for these factors can be questioned. They are likely best left as descriptive data as represented in Table X and as compared to the total frequency distribution of subgroup means in Table XV. Whether a statistical test is appropriate or not, the variation among means as presented in Table X is not particularly large. A maturity level associated with grade level did not produce a strikingly large effect on mean scores, although the range from the 9th grade average to the 12th grade average did manage to exceed the standard deviation of subgroup means.

The means which were used in the basic contrast, and which are means averaged over all seventeen replicates, are presented in Table XI.

TABLE XI

SUBGROUP MEAN ESSAY SCORES BY INSTRUCTIONAL SEQUENCE AND TIME OF TESTING, AVERAGED OVER ALL SEVENTEEN REPLICATES

	1st TESTING	2nd TESTING	3rd TESTING	
LOGIC FIRST	62.84	60.68	61.24	61.59
LOGIC SECOND	60.17	62.59	63.48	62.38
NO LOCIC	59.16	61.19	61.99	60.78
	61.02	61.49	62.24	61.58

The gain associated with logic instruction in the logic first group was a negative one, -2.16, and it was +.87 for the logic second group, making an average negative gain associated with logic of -1.29. The average gains associated with times when no logic was taught were +2.43 in the logic second group and +2.03 and +.80 in the third class, all of which averaged out to be +1.75. In a strict sense, because a one-tailed alternate hypothesis had been proposed in a positive direction, any negative result is sufficient to conclude "no effect found," but this somehow seemed like a cheap way out for investigators who admittedly were quite desirous of demonstrating that logic instruction would prove

to be a good thing. It was of some consolation, then, to find that the negative result was still well within the usually accepted bounds for chance variation. Using a pooled error term from an analysis of variance (Table XIV), the resultant t value for the contrast was -1.074, a quite probable result if the logic instruction did in fact have no effect ($p \simeq .3$). The computation of the t value is presented in Table XII.

TABLE XII

BASIC CONTRAST USING CELL MEANS OF ESSAY SCORES

Average experimental time period gains	-2.16 + .87 = -1.29
Average control time period gains	2/3(2.43 + 2.03 + .80) = +3.51
Sum of contrast coefficients squared	7.111
Pooled mean square error	47.225
Number of replications	17
Number of scores per cell per replication	1

$$t = \frac{(-1.29 - 3.51)17}{\sqrt{(47.225)(7.111)(17)}} = \frac{-81.60}{\sqrt{5708.968857}} = \frac{-81.60}{75.558} = -1.08$$

There was no discernable pattern of positive or negative contrast results according to grade level or test condition (Table XIII) that was easy to explain other than by chance variation; no grade level had more than two nor less than one positive result.

TABLE XIII

DISTRIBUTION OF CONTRAST RESULTS FROM EACH REPLICATE (TEACHER) GROUPED BY GRADE LEVEL AND TESTING CONDITION

	In-Class Writing	Out-of-Class Writing
9th	-2.60 +2.52	-8.19 +11.48
10th	-6.08 4.34 -.70	-10.44 -31.66
11th	-25.76 -20.00 +8.66	-16.38 -8.32
12th	+8.54	-5.52 +17.92

An analysis of variance of fifteen actual and one estimated replicate revealed no factor having an effect at conventional levels of significance, except for one interaction term (testing time by instructional sequence) significant at the .05 level with a pattern that had no apparent connection with the experiment. If the reasonable hypothesis that the scores would average higher for each succeeding grade level had been part of the study and preplanned, this likely would have been significant. Otherwise the subgroup means appeared to fall into a random pattern. The summary table for the analysis of variance is presented in Table XIV.

Opinions of Teachers Who Administered the Experiment

At the conclusion of the experiment, twelve experimental teachers completed a questionnaire which asked about their perception of student response, the "success" of the logic instruction, the appropriatness of the concepts, and whether or not they planned to teach the logic again. The responses to the most pertinent questions are tabled in Table XVI. The opinions of the teachers who responded certainly were more positive towards logic than implied by other data gathered. Only one stated that he "did not plan to use the logic again" compared to seven who said "yes" and four who said "maybe". Six rated the experiment a "success" and four suggested "maybe". Only one reported consistent negative pupil reaction. Were these responses biased by the fact that six teachers did not respond? Did these teachers perceive something that was not reflected in the essay ratings, implying the need for measures of different dependent variables? Or were their perceptions simply misled by the novelty of it all? Or were they too kind in their responses even though the questionnaire was completed anonymously?

On the same questionnaire eight teachers stated that their training in the workshop was adequate, two responded no. and two responded maybe. Some of the responses were accompanied by comments that a longer training session was needed

Teacher verbal reactions in informal feedbacks provided an even greater contrast to the results obtained from essay scores. A large majority were quite favorably disposed toward the nature of symbolic logic as a valid cognitive structure and its potential in composition instruction. Why, then, were they enthusiastic despite the lack of support from analysis of essay data? The teachers provided reason why the composition instruction did not have a payoff in this particular study but yet might work in future, modified trials. Inadequate time was a significant factor; inadequate time for training of teachers in symbolic logic and inadequate time for teaching the material to students. In addition, the sparsity of diverse materials with ample flexibility for individual adaptation coupled with few available ideas for varying instructional strategies to meet the needs of a spectrum of student abilities handicapped the teaching end of the experiment.

Summary of Results

1. Eighteen English teachers in grades 9 to 12 from seven school systems volunteered for the study and attended a two-week workshop, and subsequently taught symbolic logic to some of their classes. Seventeen of these teachers were able to include classes and teach logic in a way that seventeen replications of the study provided data for subsequent analysis. Three of the replications required a class taught by another teacher as a control.

2. The study produced 1264 pupil essays for scoring and analysis. Each essay was assigned a score based on nine subscores. The estimated inter-rater reliability of the average of four raters' essay scores was $r = .77$.

3. All subscores were highly intercorrelated. A principal axis factor analysis of the nine subscores of the essay ratings revealed that only one factor was involved; no other factors appeared.

4. Pupils averaged little over half correct on the 18-item multiple choice test about interpretation and analysis of sentence structure logic. The test had a overall mean of 9.42 and a standard deviation of subgroup means of 1.49 in the study. Variation among grade level means was not large, the range amounting to only 1.27 items.

5. Logic instruction had a statistically significant effect ($p < .01$ using a binomial test; $p < .05$ using a planned comparison t test) on scores on the sentence logic test. The average gain associated with logic instruction exceeded the average associated with control instruction by .62 items; this difference was approximately equivalent to 40% of a standard deviation of subgroup means.

6. The average score assigned to the essays was 61.48, and the standard deviation of subgroup means was 9.17, and the standard deviation of individual scores was 32.5. The variation among grade levels was not great and not significant according to a conservative (and probably meaningless) statistical test. There were differences, however, which were ordered in the expected direction--from low for ninth to high for twelfth grade. No difference was noted between in or out of class writing conditions.

7. Logic instruction by the teachers in the study produced no noticeable improvement on pupils' essay writing.

8. Response by twelve teachers in the study who returned a questionnaire was generally positive in regard to their plans to teach logic again (responses made prior to knowledge of results from the study), their assessment of pupil response, and the appropriateness of the concepts.

T A B L E XIV

ANALYSIS OF VARIANCE OF ESSAY TEST SCORES FOR 15 ACTUAL AND ONE ESTIMATED REPLICATE

SOURCE	d.f.	MS	F
Grade Level (G)	3	912.0	2.408
Writing Condition (W)	1	154.4	.408
G x W	3	232.5	.614
(a) Replication w/GW (R)	8	378.8	
Class W/R or Instructional Sequence (C)	2	47.2	.355
C x G	6	96.0	.722
C x W	2	55.6	.418
C x G x W	6	96.0	.721
(b) C x R	16	133.0	
Testing Time (T)	2	28.1	1.234
T x G	6	14.0	.615
T x W	2	32.8	1.443
T x G X W	6	23.4	1.030
(b) T x R	16	22.7	
C x T	4	54.7	3.300
C x T x G	12	20.4	1.230
C x T x W	4	9.0	.544
C x T x G x W	12	14.9	.901
(b) C x T x R	32	16.6	

(a) With hesitant assumption that replicates can be treated as a random factor, MS for replicates used as error term for effects listed above it in table.

(b) Part of pooled error term used for the basic contrast.

TABLE XV

GRADE LEVEL, WRITING CONDITION, AND TIME OF WRITING MEANS COMPARED WITH TOTAL DISTRIBUTION OF SUBGROUP MEANS

Grade Level Means

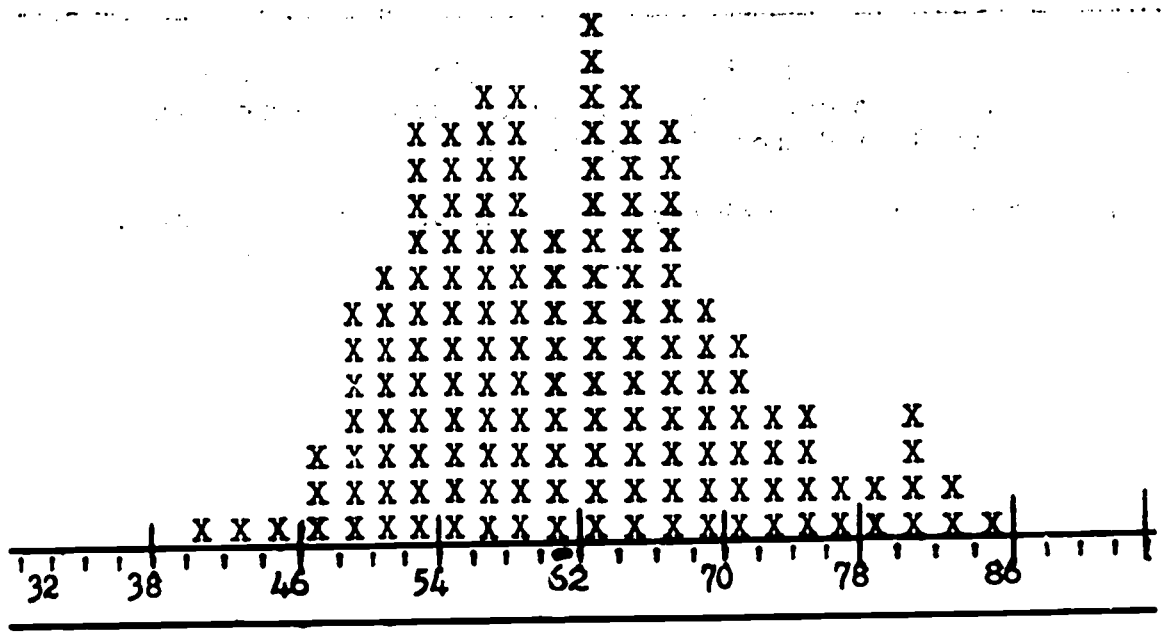
12th	66.7
11th	65.7
10th	57.8
9th	57.3

Writing Condition Means

In Class	60.8
Out of Class	62.9

Testing Time Means

1st	61.0
2nd	61.5
3rd	62.2



Distribution of Subgroup Mean

Grand Mean = 61.6

T A B L E XVI

TABULATION OF TEACHER RESPONSES ON QUESTIONNAIRE
AT CONCLUSION OF EXPERIMENT

Question	Responses	
Do you feel the logic experiment has been a success?	6	Yes
	1	No
	4	Maybe
	1	No response
How appropriate were concepts for your students?	1	Too easy
	2	Too difficult
	1	Divided between class (written comment)
	6	About right
Do you plan to continue to teach logic as part of your regular curriculum?	7	Yes
	0	Yes for some classes (written comment)
	1	No
	4	Maybe
How favorable do you feel student response has been?	4	Vary favorable
	5	Fairly favorable
	0	Indifferent
	1	Negative
	0	Downright belligerent
	2	Classes varied between favorable and unfavorable (written comment)

DISCUSSION

The major original hypothesis of improvement in pupil argumentative writing as a consequence of symbolic logic instruction was not supported by the data. Not surprisingly, in relation to the second hypothesis, pupils did show some improvement in ability to discriminate among sentences on an objective test. It appears, at a quick glance, that little remains to be said about the study. Yet, consideration of some other facts presents certain anomalies that deserve further considerations, and both the reported results and the related incongruities may have implications for the underlying rationale of a logic-oriented composition building rhetoric.

For one thing, the results do not correspond with those obtained from a controlled pilot study previously conducted by the project director (Klein, 1967). Using the same basic system of instruction and evaluation, pupils in the experimental section of the pilot study scored significantly higher than those in the control section. Some element or combination of elements such as different perceptions of time limits; different teaching in terms of knowledge presented, mannerisims, intensity, directions given, etc.; or types of pupil (unlikely since the pilot study pupils were in a class and school very similar to those in this study) that contributed to positive results in the pilot study was absent in the broader field experiment.

Secondly, the lack of obtained differences between in-class and out-of-class writing conditions was a bit of a surprise. Based on previous experience including the pilot study, it was reasoned that the very nature of symbolic logic itself coupled with the vast complexities of the writing process almost demanded the greater opportunity for thought and planning associated with a take-home rather than an in-class assignment (the in-class writing conditions factor was added to provide greater control to the study). Possibly the in-class writers did not show instruction-associated improvement because they did not have the opportunity and the out-of-class writers did not show improvement because in general they lacked "motivation" to use time outside of class, but this explanation does not square with teacher assessment of pupil response.

Third, teacher responses were quite positive while the results from the pupil essays were not. They apparently have either seen some potential in symbolic logic instruction that has not been realized in terms of pupil writing or they have been conned into the whole idea and remained conned by extraneous factors and the general positive affect of the project or they have seen attractive outcomes not measured by the study.

What speculations might help account for these possible incongruities? First of all, it should be recognized that the treatments in their field experiment represented only one specific derivation of an instructional strategy from a general rhetoric and rationale. One derivation by no means exhausts the instructional possibilities inherent in the rhetoric, and other derived strategies conceivably could have different effects. Additionally, certain factors were fixed for the

usual reasons--time limitations, specific nature of situation in co-operating schools, and resource limitation.

The specificity of the particular derivation employed as a treatment meant that positive results would provide a strong boost for the rhetoric's validity but that negative results could not definitively destroy it. Specific aspects of the study which might account for problems accompanying the findings are as follows:

1) The general, all-purpose explanation for lack of predicted results is that the measuring instruments were not sensitive enough to pick up any results that did occur. There is a craven cop-out that belatedly reveals inadequate confidence in the original operational definition. Yet, educational measurement is never quite what we wish it to be, and measurement in the field of composition is especially perilous. It could have been that results were there that were not picked up by the study's rating system; there is really no empirical evidence to claim otherwise and no way of finding out save through an exhaustive rescoring of many or all of the 1264 essays by alternative means. However, both principal investigators intuitively feel that the measuring instruments did not show predicted results because the results were not there in the first place. Moreover, even though the essay ratings are not highly reliable, the numbers of essays and raters was large enough to very likely cause results to show through a general insensitivity if in fact the insensitivity was largely brought about by unreliability.

2) Symbolic logic is a highly sophisticated field of study requiring a high degree of self discipline and intellectual application. Practice is important; perhaps even more important than investigators assumed when organizing the study. It could be that a two-week concentrated training session in symbolic logic was neither ample nor the best way to organize such training. Little time was given teachers for experimentation, practice and application on their own. Any additional attempts in this area certainly should be based upon an expanded and less concentrated study of the material.

3) Class time allowed for teaching of the units ranged from four to six weeks. In view of teacher reactions, it appears that such a length of time leaves little opportunity for attempts at practical applicaotion. In view of findings such as the McColly and Remstad study,

> Until more is known, the best inference seems to be that to teach composition effectively, teachers should give a weekly writing task on which they base about two and a half days of practical explanation, student practice, discussion, revising, rewriting, etc.

it seems that the assumed length of time which was previously considered functional was in fact inadequate. Students in the lower secondary grades appeared especially to need more time although the entire range of grades involved appeared to have difficulties. Low scores on both essays and objective tests tend to bear this out.

Because of time limitations, important related introductory units in inductive reasoning and vindication models were deleted and hence potentially important aspects of symbolic logic were not included. Indeed it is quite possible that the importance of induction as a process or as a process in relation to deduction has been underestimated and deserves further exploration.

4) Too, one must consider the potential of certain other kinds of cognitive elements which might play a significant role—argumentation models premised upon inductive and/or deductive frameworks, vindication-validation models for analysis, etc. Such an approach might broaden the range of the study and hence allow more flexibility in factor relationship possibilities.

5) One of the most obvious problems in the study arose from a limited amount of student material available. Anticipated opportunities for teacher produced materials did not materialize and most teachers relied upon mimeographed handouts or slight modifications of materials used in the teacher training workshop. Especially hard hit were slow learners where both time and material limitations worked a hardship. Development of additional materials geared to slow learners along with options in instructional strategies which might be used could be a step toward needed refinement.

In the opinion of the investigators, then, possibilities still exist for an effective composition rhetoric based on elements from symbolic logic. This study failed to demonstrate such possibilities. If any possibilities are to be demonstrated, a different research approach seems called for. In this case, the gamble to go from a pilot study and informal experience to a field experiment entailing a specific implementation of the rhetoric was a losing one. Further study might be conducted by working out several instructional strategies and materials sets, and comparing them in a controlled experiment. This, however, appears to be uneconomical and not fessible on the basis of what we how know. It is suggested that a better avenue now would be to go back to a smaller, more carefully-monitorial descriptive approach. With careful attention paid to seemingly relevant differences in pupil characteristics, one or a few teachers could try to teach symbolic logic the best way they could and stick with it until unmistakeable change occurs in pupil writing (or until the teachers are completely exhausted). A careful description of this experience could then be examined, and from it a sort of a model might be formulated—one that would desirably provide instructional and cognitive flexibility according to pupil progress and intermediate responses to instruction. This model could then again be tested in a wider and perhaps more flexible field situation.

The price of this research effort in time if not in money promises to be high, but if composition is as important as many believe and if the conceptual domain of symbolic logic has the rich potential that its adherents perceive, the risk is more than acceptable.

BIBLIOGRAPHY

Bateman, D. & Zidonis F. The Effects of a Study of Transformational Grammar on the Writing of Ninth and Tenth Graders.
Champaign: NCTE, 1966.

Beier, E. G. et. al. A Perception of Grammaticalness.
1967.

Blake, R. W. The Effect of Special Instruction on the Ability of Seventh and Eighth-Grade Pupils to Write Composition and Understand Poety and Short Fiction.
N. Y.: State University College, 1966.

Bone, N. L. A Comparison Between Objective and Subjective Methods in Evaluating Stud, Writing.
Unpublished Master's Thesis, University of Kansas, 1964.

Bormuth, J. R. "Readability: A New Approach", Reading Research Quarterly, 1, 1966.

Bormuth, J. R. Relationship Between Selected Language Variables and Comprehension Ability and Difficulty.
1964

Braddock, R., Lloyd-Jones, R., & Schoer, L. Research in Written Composition.
National Council of Teachers of English, 1963.

Budd, W. C. "Research Designs of Potential Value in Investigating Problems in English," Research in the Teaching of English, 1 Spring, 1967.

Burton, D. L. & Arnold, L. V. Effects of Frequency of Writing and Intensity of Teachers Evaluation Upon High School Students' Performance in Written Composition.
U.S. Office of Education Cooperative Research Project No. 1523, 1963.

Buxton, E. W. An Experiment to Test the Effects of Writing Frequency and Guided Practice Upon Students' Skill in Written Expression.
Unpublished doctoral diss., Stanford University, 1958.

Christiansen, M. A. The Relative Effectiveness of Two Methods of Teaching Composition in Freshman English at Metropolitan Junior College.
Doctoral Thesis, University of Kansas, 1964.

Crockett, W. Effects of Age, Social Status, Sex, and Race Upon the Understanding of Work Meanings Independent of Sentence Context.
1966.

BIBLIOGRAPHY (CONTINUED)

Diederich, P. B. "How to Measure Growth in Writing Ability," English Journal, 55, 1966.

Dressel, P., Schmid, J. & Kincaid, G. "The Effect of Writing Frequency Upon Essay-Type Writing Proficiency at the College Level." Journal of Ed. Res., 46, Dec., 1952, pp. 283-293.

Fellers, A. L. Problems in Writing in College Composition Classes. Unpublished doctoral dissertation, Stanford University, 1953.

Follman, J. C. & Anderson, J. A. "An Investigation of the Reliability of Five Procedures for Grading English Themes." Research in the Teaching of English, 2, Fall, 1967.

Godshalk, F. et. al. The Measurement of Writing Ability. New York: College Entrance Examination Board, 1966.

Heys, F. Jr. "The Theme-A-Week Assumption: A Report of an Experiment." English Journal, 51, May, 1962, pp. 320-322.

Hunt, K. W. Sentence Structure Used by Superior Students in Grades Four & Twelve, and by Superior Adults. U.S. Office of Education Cooperative Research Project No. 5-0313,1966.

Kahler, A. D. The Effects of Programmed Grammar and Journal Writing Ability: An Exploratory Study. Unpublished doctoral dissertation, Oklahoma State University, 1965.

Klein, M. L. An Experiment in the Use of Symbolic Logic As a Factor in Exposition. Cooperative Curriculum Development Center,1967,

Klein, M. L. Symbolic Logic and Its Application to Writing Cooperative Curriculum Development Center, 1967.

Lokke, V. L. & Wykoff, G. S. "Double Writing in Freshman Composition: Experiment," School and Society. 68, Dec., 1948, pp. 437-439.

McColley, J. Effects of a Method of Teaching Sentence Structure Upon Sentence Structure Used in Writing. U.S. Office of Education Cooperative Research Project No. 2-092. 1965.

McColly, W. & Remstad, L. Comparative Effectiveness of Composition Skills Learning Activities in the Secondary School. U.S. Office of Education Cooperative Research Project No. 1528, 1963

Mellon, John. Transformational Sentence Combining: A Method for Enhancing the Development of Syntactic Fluency in English Composition, NCTE,

BIBLIOGRAPHY (CONTINUED)

O'Donnell, L. C. The Correlation of Awareness of Structural Relationships in English and Ability in Written Composition.

Pitt. Observations: 1965 Institute.
Unpublished collation, Fresno State College, 1966.

Sam, N. H. & Stine, E. S. Structural Analysis of the Written Composition of Intermediate Grade Children.
Lehigh University, 1965.

Scannell, D. P. & Haugh, O. M. Teaching Composition Skills With Weekly Multiple Choice Tests in Lieu of Theme Writing, 1968.

Sherwin, J. Stephen. Four Problems in Teaching English: A Critique of Research.
Scranton: International Textbook Co., 1969.

Sutton, J. T. & Allen, E. D. The Effect of Practice and Evaluation On Improvement in Written Composition.
U.S. Office of Education Cooperative Research Project No. 1993. 1964.

Tillman, M.H. Level of Abstraction in the Written Compositions of Children Varying in Intelligence and Age. 1967.

Tovatt, A. L. & Miller, E. L. Oral-Aural-Visual Stimuli Approach to Teaching Written Composition to 9th Grade Students.
Ball State University, 1967.

Wolf, M. H. Effect of Writing Frequency Upon Proficiency in a College English Course.
U.S. Office of Education Cooperative Research Project No. 2846, 1966.

West, William W. "Written Composition." Chapter V in "Language and Fine Arts." Review of Educational Research.
37:2, April, 1967.

Zoellner, Robert. "A Behavioral Approach to Writing."
College English, January, 1969.

APPENDIX

APPENDIX A

Logic Test

I) Fill in the space on the answer sheet for the conclusion that can be drawn from the argument in each case:

1) If you study, you will pass. You didn't pass, therefore:
   A) you studied.
   B) you didn't study.
   C) it's not true that you didn't study.
   D) you might not have studied.

2) If it rains, the crops will grow. If the crops grow, we'll make money, therefore:
   A) we'll make money.
   B) we might not make money.
   C) if it rains, we'll make money.
   D) if we made money, it rained.

3) Either we improve our internal problems or we'll lose many friends abroad. And, if we should lose many friends abroad, our status as a nation will be greatly reduced. We aren't improving our internal problems. Hence:
   A) We'll not lose many friends abroad.
   B) Our status as a nation will be greatly reduced.
   C) Our status as a nation will not be greatly reduced.
   D) Our status as a nation might not be greatly reduced.

II) Fill in the space on the answer sheet for each of the following that has the conclusion underlined:

4) A) Joe is failing math, so I'll have to help him.
   B) Joe is failing math, so I'll have to help him.

5) A) All of Ernest Hemingway's books are good. This book is by Hemingway, so it must be good.
   B) All of Ernest Hemingway's books are good. This book is by Hemingway, so it must be good.
   C) All of Ernest Hemingway's books are good. This book is by Hemingway, so it must be good.

6) A) Since the weather is bad, the picnic is called off.
   B) Since the weather is bad, the picnic is called off.

7) A) We won't win because our star pitcher isn't here.
   B) We won't win, because our star pitcher isn't here.

III) Fill in the space on the answer sheet for the letter/s that represent/s the correct translation/s in each of the following:

8) "Of a good beginning cometh a good end."
   A) If you don't have a good beginning, you'll not have a good end.

B) If you had a good end, you had a good beginning.
C) If you have a good beginning, you'll have a good end.
D) All good things must come to an end.

9) Courage is enough for winning.
A) If it's not the case that you don't have courage, you'll not win.
B) If you won, you have courage.
C) If you don't have courage, you won't win.
D) If you have courage, you'll win.

10) "To be great is to be misunderstood"
A) If you're misunderstood, you're great.
B) If you're great, your're misunderstood.
C) If you're not misunderstood it's not the case that you're not great.
D) If it's not the case that you're not great, then you're not misunderstood.

11) "Old women should not seek to be perfumed"
A) If you're perfumed, you're an old woman.
B) If you're not an old woman, you should seek to be perfumed.
C) If you're an old woman, it's not the case that you should not seek to be perfumed.
D) If you're an old woman, you should not seek to be perfumed.

12) "Anyone who conducts an argument by appealing to authority is not using his intelligence; he is just using his memory."
A) If you have no memory, you have no intelligence.
B) If one conducts an argument by appealing to authority, he is not using intelligence, just his memory.
C) If you don't use your memory, you can't conduct an argument.
D) If you arn't conducting an argument, you aren't using your memory.

IV) Fill in the space on the answer sheet for the letter/s of those statements that represent/s correct translation/s of the numbered statements.

13) If the plane crashes, it will burn.
A) If the plane burns, it has crashed.
B) If the plane doesn't burn, it hasn't crashed.
C) Either the plane crashes or it burns.
D) Either the plane doesn't crash or it burns.

14) If we don't leave quickly, we'll be covered by the avalanche.
A) Either we aren't covered by the avalanche or we don't leave quickly.
B) Unless we leave quickly, we'll be covered by the avalanche.
C) If we aren't covered by the avalanche, then it's not the case that we didn't leave quickly.

15) It's not the case that either Joe isn't going or Harry is going.
   A) Either Joe is going or Harry isn't
   B) If Joe is going, Harry is going.
   C) It's not the case that Joe isn't going and at the same time Harry isn't.
   D) All of the above.

16) If Mary isn't sad, Alfred has not left home.
   A) It's not the case that Alfred leaves home and Mary is not sad.
   B) Either Alfred doesn't leave home or Mary will be sad.
   C) If Alfred leaves home, Mary will be sad.
   D) Either Alfred doesn't leave home or it's not the case that Mary will not be sad.

V) Fill in the space on the answer sheet for the letter of the valid conclusion for each of the following arguments:

17) If we continue to escalate the war in Viet Nam, we will further alienate the neutral countries of southeast Asia. But only if we do not alienate the neutral countries can we halt the spread of Communism into southeast Asia. But it cannot be that Communism will continue to spread in southeast Asia and that we can win the battle against Communism in that area. But what we want is to win the battle against Communism in southeast Asia. Therefore:
   A) We should continue to escalate the war in Viet Nam.
   B) We should not continue to escalate the war in Viet Nam.
   C) Neither of the above.

18) Only if we are able to stop the flow of gold from the United States to foreign markets will we have a sound dollar. But unless we are able to balance the budget we will not be able to stop the flow of gold abroad. We cannot both balance the budget and continue to expand our federal programs. Therefore:
   A. We will have a sound dollar.
   B. We will not have a sound dollar.
   C. If we wish to have a sound dollar, we must expand our federal programs.
   D. If we continue to expand our federal programs, we will not have a sound dollar.

APPENDIX B

Directions for Essay Writing

Choose a current topic, about which you have personal convictions, and write a paper designed to persuade the reader through careful reasoning that your view is a sound one.

It is suggested that this paper should probably not be longer than 1500 words.

The following topics are suggested only to give some idea of the nature of the assignment. If you wish to write on one of them, you may, but it isn't necessary.

1. We should Get Out of Vietnam!
2. We Should Have an Income Tax Increase.
3. The School Year Should Be Extended to 10 Months.
4. The Government Should Force Water Pollution Controls.
5. E. Hemingway Fails to Achieve His Purpose in *A Farewell to Arms*.
6. Most Teen-agers Have Too Much Money to Spend These Days.

APPENDIX C

Guidelines for Evaluation of Essays

GENERAL COMMENTS:

We are primarily concerned with logic and not rhetoric insofar as they can be separated. In addition, mechanics of writing should play little, if any, role.

Please remember that the operating hypothesis is that a study of symbolic logic improves sentence structure and organization of the paper. Although rhetorical factors are important, we are not primarily concerned with them here.

Following is an itemized elaboration of elements to be considered.

SENTENCE STRUCTURE:

1. Clarity-
This item is largely self-explanatory. Does each sentence tell in an understandable way? Is the reader forced to ponder the meaning? Is the sentence ambiguous? Incoherent?

2. Structural variety-
Remember, logical sentences are usually in the form of assertions. Prescriptions, directives, and questions should have little role in the paper.

Logical assertion types that often appear in arguments include the following:

a. "If—then" sentences
b. "Either—or" sentences
c. "Neither—nor" sentences
d. "X and Y" sentences

The first two types are often avoided by writers, but can be helpful for the effective argument.

3. Logical effectiveness-
Does the sentence make a reasoning impact? Does it reveal logical insight? Is the writer willing to take a stand, or does he prefer to "play it safe"?

4. Rhetorical force-
Is the writing style mature? (Close ties to no. 2) Is the vocabulary in keeping with the intellectual tone of the paper?

In addition, is there variety in sentence patterning?
i.e. subordinate clause + independent clause
introductory sentence modifers + indep. clause,
etc.

ORGANIZATION:

1. Soundness-
   Does the paper reach a valid conclusion based on true assertions? Just because you disagree with the conclusion is not a necessary indication of the paper's inadequacy. Remember too that even if backing for the assertions of the paper is inadequate, the conclusion could still be valid and the number rating given should reflect credit for this validity. Don't mistake an array of facts for solid assertions, but do remember that empiric support is often helpful.

2. Sophistication-
   This refers only to the maturity, originality, and complexity of the argument and not to the eloquence of the paper. A sophisticated argument is one that reflects mature intelligent thinking.

3. Unity-
   Does the paper move well or does the writing appear disjointed? Does one paragraph flow into another? Is the relationship between backing assertions clear?

4. Rhetorical force-
   Again, a matter of style. Does the paper possess more than just logical soundness? Please don't confuse with eloquence. A simple straightforward style is usually more persuasive than Victorianized embellishments.

5. Total effectiveness of paper-
   What is the overall impact of the paper?

FOLLOWING IS A PAPER WRITTEN BY A HIGH SCHOOL STUDENT AND AN EVALUATION WITH EXPLANATORY COMMENTS RELATIVE TO THE RATINGS.

THE RUNNING OUT

It is time for the Chicago Bears to recognize the predicament facing them; they either find a top notch quarterback to replace Rudy Bukich or prepare themselves for another second division finish.

A pro football team today must have a good steady quarterback in order for that team to be a success. Bart Starr makes the Packers go, and the Baltimore Colts would be lost without John Unitas. These are good quarterbacks; Rudy Bukich isn't.

For some reason, George Halas, owner-coach of the Bears, doesn't seem to realize this. He isn't even trying to replace Bukich. He hasn't tried to trade for a good quarterback, and he hasn't picked one in the college draft for the past four years; that is, one worth mentioning. Larry Rakestraw of Georgia, the one he drafted four years ago has never lived up to expectations, and I doubt if he ever will.

It's obvious Halas plans to use Bukich as his regular quarterback again next season. So, Bears, prepare yourselves to dwell in the second division during 1967.

SENTENCE STRUCTURE:

1. Clarity-
   Sentences are concise and clear. There is no befuddlement or ambiguity. One may not agree with the sentences, but one has little difficulty understanding what is meant.

2. Structural variety-
   The assertion dominates. A rather unique use of an "either-or" type is in paragraph one. Structural variety enhanced by clever use of semi-colon not often found in high school writing.

3. Logical effectiveness-
   The sentences carry the argument well. They reveal insights but are somewhat lacking in depth.
   In any event, I could certainly understand why an evaluator would score this item and the previous one with a 10.

4. Rhetorical force-
   The sentences are forceful. Again, use of semi-colon and subordinate elements indicates rhetorical maturity.

ORGANIZATION:

1. Soundness-
This paper is valid. The writer employs what logic refers to as a Disjunctive Syllogism.

Either X or Y is the case.
X is not the case.
Therefore, Y must be the case.

Either Bears find a quarterback or prepare for a second division finish.
They aren't finding one.
Therefore, they'll finish in second division.

However, depth of support for the basic premises is weak. If your view suggests that a good team is feasible with only a mediocre quarterback, then you will want to examine closely his support for the opening assertion. If you think Bukich is a good quarterback, then you'll want to question support for his second assertion.

2. Sophistication-
The argument is a basic valid form. It reveals logical insight. It could, however, be enhanced with a more subtle valid form.

3. Unity-
The paper moves well. Each paragraph relates to the preceeding and following. There is a sence of cohesion.

4. Rhetorical force-
This seems fixed by that of the individual sentence, the impact of which relates well to the entire paper.

5. Total effectiveness-
Valid argument. Mature sentences. Logical assertions. Lacking in subtlety. Lacking in depth of assertions support.

PAPER NUMBER_____ GRADER_____

SENTENCE STRUCTURE:	LOW		MIDDLE		HIGH
1. Clarity	2	4	6	8	(10)
2. Structural variety	2	4	6	(8)	10
3. Logical effectiveness	2	4	6	(8)	10
4. Rhetorical force	1	2	3	4	(5)

_____

ORGANIZATION:					
1. Soundness	5	10	15	(20)	25
2. Sophistication	2	4	6	(8)	10
3. Unity (effective transition, etc.)	2	4	6	8	(10)
4. Rhetorical force	2	4	6	(8)	10
5. Total effectiveness of paper	2	4	6	(8)	10

_____

TOTAL SCORE _____

APPENDIX D

Test Scores for each Subgroup in each Replicate

11th grade
1st testing condition
teacher #14

Time of Testing 1	2	3	Sequence (Class)
9 / 7.67	9 / 11.11	8 / 12.25	1
10 / 6.50	10 / 8.40	9 / 10.56	2
8 / 8.88	9 / 8.56	8 / 8.43	3

Contrast = +4.63

11th grade
2nd testing condition
teacher #11

1	2	3	
10 / 10.50	10 / 9.90	8 / 10.88	1
9 / 11.00	8 / 10.[illegible]	8 / 11.25	2
9 / 6.22	8 / 9.13	8 / 9.50	3

Contrast = +1.48

11th grade
2nd testing condition
teacher #12

1	2	3	
6 / 11.83	5 / 13.60	4 / 14.75	1
3 / 9.67	6 / 12.33	6 / 13.83	2
8 / 8.25	8 / 6.38	8 / 10.38	3

Contrast = +.08

Number of pupils per subgroup indicated in upper right corner of each cell.

Test Scores for Each Subgroup in Each Replicate (Continued)

9th grade
1st testing condition
teacher #4

Time of Testing 1	2	3	
10 / 6.73	8 / 8.38	10 / 8.80	1
8 / 10.13	8 / 8.13	7 / 9.67	2
8 / 9.25	10 / 8.50	9 / 7.33	3

Contrast = +5.80

9th grade
1st testing condition
teacher #18

1	2	3	
12 / 9.33	10 / 9.50	8 / 8.88	1
8 / 11.00	9 / 8.89	7 / 9.43	2
8 / 9.63	10 / 7.50	8 / 9.25	3

Contrast = +2.70

10th grade
1st testing condition
teacher #7

1	2	3	
4 / 11.75	4 / 10.00	4 / 9.25	1
6 / 9.00	6 / 10.17	6 / 10.33	2
4 / 9.25	4 / 10.25	3 / 9.33	3

Contrast = -2.42

Test Scores for Each Subgroup in Each Replicate (Continued)

10th grade
2st testing condition
teacher #5

Time of Testing

1	2	3	Sequence (Class)
5 / 8.00	6 / 8.00	6 / 8.67	1
6 / 5.83	6 / 8.33	6 / 11.17	2
6 / 6.67	6 / 9.00	6 / 7.67	3

Contrast = +.51

10th grade
1st testing condition
teacher #9

Time of Testing

1	2	3	
8 / 8.88	8 / 8.50	8 / 9.38	1
10 / 9.60	10 / 9.10	9 / 10.56	2
10 / 6.80	10 / 7.70	10 / 8.00	3

Contrast = +.61

9th grade
2nd testing condition
teacher #2

1	2	3	
7 / 8.00	7 / 10.29	6 / 10.50	1
9 / 10.00	8 / 7.63	10 / 10.60	2
10 / 8.00	8 / 10.13	10 / 6.50	3

Contrast = +4.44

est Scores for Each Subgroup in Each Replicate (Continued)

9th grade
2nd testing condition
teacher #3

Time of Testing 1	2	3	
7 / 9.00	8 / 9.00	8 / 8.38	1
8 / 6.63	8 / 8.13	8 / 9.00	2
7 / 7.71	8 / 8.50	7 / 6.43	3

Contrast = +.72

10th grade
1st testing condition
teacher #6

1	2	3	
6 / 11.00	6 / 9.50	6 / 10.83	1
5 / 5.60	4 / 9.00	7 / 8.14	2
6 / 6.33	6 / 8.33	7 / 10.00	3

Contrast = -7.07

12th grade
1st testing condition
teacher #15

1	2	3	
10 / 8.40	9 / 11.33	10 / 8.20	1
9 / 8.67	11 / 9.55	7 / 9.29	2
9 / 8.56	11 / 9.55	9 / 9.44	3

Contrast = +1.50

Test Scores for Each Subgroup in Each Replicate (Continued)

12th grade
2nd testing condition
teacher #16

1	2	3	
8 / 9.25	11 / 11.73	10 / 11.60	1
9 / 11.00	9 / 11.56	9 / 11.11	2
6 / 8.33	9 / 9.59	9 / 10.67	3

Contrast = +10

12th grade
2nd testing condition
teacher #17

1	2	3	
9 / 7.78	7 / 10.43	7 / 9.71	1
10 / 10.40	9 / 10.00	9 / 11.89	2
8 / 8.63	9 / 9.56	8 / 8.86	3

Contrast = +4.67

10th grade
2nd testing condition
teacher #8

1	2	3	
8 / 9.50	8 / 9.50	7 / 9.86	1
5 / 8.00	6 / 10.83	5 / 9.40	2
6 / 11.00	6 / 9.83	6 / 11.00	3

Contrast = -2.78

Test Scores for Each Subgroup in Each Replicate (Continued)

10th grade
1st testing condition
teacher #10

Time of Testing 1	2	3	
9 / 8.67	6 / 9.00	7 / 11.00	1
10 / 9.20	10 / 9.80	9 / 11.78	2
9 / 9.56	9 / 7.89	9 / 10.33	3

Contrast = +1.40

11th grade
1st testing condition
teacher #13

1	2	3	
9 / 7.33	10 / 10.60	10 / 10.00	1
10 / 9.00	10 / 9.50	10 / 10.50	2
9 / 8.67	10 / 9.50	9 / 9.67	3

Contrast = +3.27

Estimated Dummy Replicate Data for Analysis of Variance Purposes (Test Scores)

12th grade
1st testing condition
teacher #0

1	2	3	
9.29	10.27	10.33	1
9.15	9.74	10.67	2
8.71	9.03	9.17	3

APPENDIX E

Essay Scores for Each Subgroup in Each Replicate

Time of Testing

9th grade
1st testing condition
teacher #4

1	2	3	
11 / 49.73	8 / 56.66	10 / 53.55	1 Logic First
8 / 57.47	8 / 63.53	7 / 61.82	2 Logic Second
8 / 49.19	10 / 54.38	9 / 54.86	3 No Logic

Contrast = -2.60

9th grade
1st testing condition
teacher #18

1	2	3	
12 / 55.63	10 / 52.38	8 / 54.41	1
8 / 61.75	9 / 60.78	7 / 67.57	2
8 / 62.72	10 / 59.72	8 / 65.22	3

Contrast = +2.52

9th grade
2nd testing condition
teacher #2

1	2	3	
7 / 61.32	7 / 54.46	6 / 53.35	1
9 / 64.78	8 / 63.94	10 / 61.97	2
10 / 50.57	8 / 48.19	10 / 50.45	3

Contrast = -8.19

Number of pupil essays per subgroup indicated in upper right corner of each cell.

Essay Scores for Each Sub roup in Each Replicate (Continued)

9th grade
2nd testing condition
teacher #3

Sequence (Class) \ Time of Testing	1	2	3
1	7 / 55.50	8 / 57.88	8 / 56.34
2	8 / 48.00	8 / 50.16	8 / 52.97
3	7 / 68.71	8 / 65.28	7 / 57.11

Contrast = +11.48

12th grade
1st testing condition
teacher #15

Sequence (Class) \ Time of Testing	1	2	3
1	10 / 62.15	9 / 65.33	10 / 53.13
2	9 / 53.22	11 / 56.11	7 / 67.00
3	9 / 67.83	11 / 67.14	9 / 73.22

Contrast = +8.54

10th grade
1st testing condition
teacher #6

Sequence (Class) \ Time of Testing	1	2	3
1	6 / 56.42	6 / 55.75	6 / 59.58
2	5 / 56.05	4 / 51.56	7 / 50.29
3	6 / 56.04	6 / 53.13	7 / 58.68

Contrast = -.71

Essay Scores for Each Subgroup in Each Replicate (Continued)

10th grade
1st testing condition
teacher #7

Time of Testing 1	2	3	
4 / 74.88	4 / 52.69	4 / 64.31	1
6 / 63.31	6 / 54.38	6 / 67.25	2
4 / 62.44	4 / 63.44	3 / 66.50	3

Contrast = -6.07

10th grade
1st testing condition
teacher #9

1	2	3	
8 / 55.03	8 / 59.13	8 / 58.94	1
10 / 59.82	10 / 60.22	9 / 64.03	2
10 / 44.55	10 / 48.30	10 / 49.50	3

Contrast = +4.34

11th grade
1st testing condition
teacher #13

1	2	3	
9 / 63.03	10 / 63.15	10 / 71.40	1
10 / 60.10	10 / 71.22	10 / 58.77	2
9 / 65.36	10 / 71.47	9 / 66.64	3

Contrast = -19.99

Essay Scores for Each Subgroup in Each Replicate (Continued)

11th grade
1st testing condition
teacher #14

Time of Testing 1	2	3	
9 / 48.72	9 / 55.47	8 / 53.88	1
10 / 53.07	10 / 61.67	9 / 59.08	2
8 / 53.06	9 / 58.94	8 / 62.72	3

Contrast = +8.65

11th grade
2nd testing condition
teacher #11

1	2	3	
10 / 73.92	10 / 65.60	8 / 62.84	1
9 / 62.86	8 / 67.91	8 / 67.31	2
9 / 65.61	8 / 68.06	8 / 71.75	3

Contrast = -16.38

12th grade
2nd testing condition
teacher #5

1	2	3	
5 / 46.50	6 / 42.21	6 / 51.00	1
6 / 41.17	6 / 46.42	6 / 47.25	2
6 / 51.79	6 / 56.46	6 / 57.00	3

Contrast = -10.43

Essay Scores for Each Subgroup in Each Replicate (Continued)

10th grade
2nd testing condition
teacher #8

Time of Testing 1	2	3	
8 / 76.56	8 / 61.63	7 / 58.11	1
5 / 56.00	6 / 67.25	5 / 56.90	2
6 / 66.96	6 / 65.08	6 / 65.29	3

Contrast = -31.67

11th grade
1st testing condition
teacher #10

1	2	3	
9 / 68.92	8 / 69.46	7 / 68.04	1
10 / 68.13	10 / 83.35	9 / 71.47	2
9 / 58.31	9 / 61.94	9 / 64.72	3

Contrast = -25.76

11th grade
2nd testing condition
teacher #12

1	2	3	
6 / 82.67	5 / 80.65	4 / 80.75	1
3 / 80.17	6 / 85.71	6 / 78.83	2
8 / 58.66	8 / 49.72	8 / 52.25	3

Contrast = -8.32

Essay Scores for Each Subgroup in Each Replicate (Continued)

12th grade
2nd testing condition
teacher #16

Time of Testing 1	2	3	
8 / 73.34	11 / 75.50	10 / 75.38	1
9 / 73.97	9 / 75.00	9 / 76.44	2
6 / 68.08	9 / 81.89	9 / 79.22	3

Contrast = -4.51

12th grade
2nd testing condition
teacher #17

1	2	3	
9 / 63.97	7 / 62.96	7 / 66.07	1
10 / 63.00	9 / 54.88	9 / 70.28	2
8 / 55.81	9 / 67.14	8 / 58.63	3

Contrast = +17.92

Estimated Dummy replicate data for analysis of variance purposes (Essay score)

12th grade
1st testing condition
teacher #0

1	2	3	
67.53	65.37	65.93	1
65.74	67.28	68.17	2
63.85	65.88	66.88	3

Sequence (Class)