



Criteria for the evaluation of high school English composition
by Donald Raymond Fostvedt

A thesis submitted to the Graduate Faculty in partial fulfillment of the requirements for the degree of
DOCTOR OF EDUCATION

Montana State University

© Copyright by Donald Raymond Fostvedt (1963)

Abstract:

The belief that valid and reliable criteria for the evaluation of English composition at the twelfth grade level could be established led to this investigation. An examination of literature revealed that during the first two decades of the twentieth century emphasis was placed upon comparison scales. In the attempts to establish criteria after 1920, numerous criteria were presented by different studies. In the studies where the criteria were tested for reliability and validity, the correlation coefficients were found to be low.

Criteria in this study were selected from nine sources in literature which were published from 1951 to 1962 by The National Council of Teachers of English, state English associations, national testing services, and individuals of national importance in the field of English. The criteria were ranked in terms of importance by nine college and 22 high school experts. Ranking was tested at the one per cent level for validity by Kendall's coefficient of concordance, "W," and agreement was found.

The ranking was as follows: (1) Development of Ideas, (2) Coherence and Logic, (3) Organization Through Sentence Structure and Paragraphing, (4) Diction, and (5) Emphasis. To determine reliability of the criteria, 30 high school English experts from 26 different high schools rated, according to each criterion, 20 themes which were selected from 256 written by the seniors of four Montana high schools. Fourteen tests were made to test at the five per cent level the ratings given each theme according to each criterion by each of the 30 experts by using an analysis of variance test and chi square tests. A chi square test was also used to test at the five per cent level the overall grades given the themes. The homogeneity of the mean grades that teachers give each theme, of the grades that students received on themes, and of the total markings given by the experts was tested. Each criterion was tested separately for a group of 20 experts and a second group of 10 experts whose ratings were received about a month after the first ratings. Agreement was found for "Organization Through Sentence Structure and Paragraphing" for the group of 20 experts and for "Emphasis" for the group of 10 experts. Since this was the only agreement, reliability of the criteria for the evaluation of English composition was not established.

Validity of the criteria for the evaluation of twelfth grade English composition was established by agreement in the ranking in terms of importance by college and high school English experts.

Perhaps the only conclusion justified by this study is that, although teachers of English composition may feel that criteria are important in evaluating themes, there is no evidence of consistency in the employment of such criteria.

CRITERIA FOR THE EVALUATION OF HIGH
SCHOOL ENGLISH COMPOSITION

by

DONALD RAYMOND FOSTVEDT

A thesis submitted to the Graduate Faculty in partial
fulfillment of the requirements for the degree

of

DOCTOR OF EDUCATION

Approved:

Wilford Franks
Head, Major Department

Wilford Franks
Chairman, Examining Committee

James O. Smith
Dean, Graduate Division

MONTANA STATE COLLEGE
Bozeman, Montana

September, 1963

ACKNOWLEDGMENT

The end results, complete with summarizations and conclusion, of a work such as this depended upon the concern and cooperative assistance of hundreds of persons who so generously devoted their time and energy.

This humble investigator wishes to take this opportunity to say "I thank you" to the 30 Montana high school English instructors, to the 256 English students, to the administrators and English teachers of the schools where the themes used in this study were written, to the college and university experts from various institutions throughout the United States, and to the members of my committee at Montana State College whose guidance was invaluable. Without their help this study would not have been possible.

D.R.F.

TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION	1
Statement of the Problem	2
Procedures	2
Selection of Criteria.	2
Establishment of Validity.	2
Establishment of Reliability	3
Limitations and Controls	4
II. REVIEW OF LITERATURE	6
Attempts at Developing a Scale for the Evaluation of Compositions, 1903 - 1950.	7
Scale Developments, 1951 - 1962.	12
Summary.	36
III. SELECTION AND VALIDATION OF CRITERIA FOR THE EVALUATION OF ENGLISH COMPOSITION	39
Selection of Criteria for the Evaluation Scale	40
Validation of Criteria by College English Experts.	44
Selection of College Experts	44
Statistical Treatment of Numerical Rankings by College Experts.	47
Validation of Criteria by High School English Experts.	49
Selection of High School Experts	49
Statistical Treatment of the Numerical Rankings by High School Experts.	50
Summary.	51
IV. ESTABLISHMENT OF RELIABILITY OF THE CRITERIA FOR THE EVALUATION OF ENGLISH COMPOSITION.	53
Selection of Themes Written by High School Students.	54

Chapter	Page
Instructions for Writing Themes	54
How Themes Were Selected	55
Establishment of Reliability of the Criteria by High School English Experts	57
Statistical Treatment of the Numerical Ratings of the High School English Experts.	58
Summary.	68
V. SUMMARY, CONCLUSION, AND RECOMMENDATIONS	71
Summary.	72
Conclusion	74
Recommendations.	75
LITERATURE CONSULTED	76
Appendix A: Statistical Computations.	81
Appendix B: Materials Pertinent to the College Expert Ratings of the Scale for Grading English Compositions.	105
Appendix C: Materials Directed to High School Experts Pertinent to the Theme Judging.	115

LIST OF TABLES

Table	Page
1. Emphasis Given Five Criteria in the Nine Sources	42
2. Rank Given by Each of the Nine College English Experts for Importance of Each Criterion	48
3. Rank Given by Each of the 22 High School English Experts for Importance of Each Criterion	51
4. Values of Chi Square with 38 Degrees of Freedom for Each of the Criteria for the First 20 Experts	64
5. Values of Chi Square with 18 Degrees of Freedom for Each of the Criteria for the 10 Additional Experts.	65
6. Cumulative Scores of Teacher's Grading on Each Theme for Combined Criteria.	83
7. Analysis of Variance--F Ratios Calculated From Sum of Squares, Degrees of Freedom, and Mean Squares.	84
8. Frequency of Overall Grades Given by 20 High School English Experts Marking 20 Themes.	88
9. Frequency of Overall Grades Given by the Last 10 High School English Experts Marking 20 Themes	89
10. Frequency of Grades for the First 20 Experts When They Graded Criterion Number One--"Coherence and Logic"	90
11. Frequency of Grades for the First 20 Experts When They Graded Criterion Number Two--"Development of Ideas"	91
12. Frequency of Grades for the First 20 Experts When They Graded Criterion Number Three--"Diction"	92
13. Frequency of Grades for the First 20 Experts When They Graded Criterion Number Four--"Emphasis"	93
14. Frequency of Grades for the First 20 Experts When They Graded Criterion Number Five--"Organization Through Sentence Structure and Paragraphing.	94

Table	Page
15. Frequency of Grades for the Last 10 Experts When They Graded Criterion Number One--"Coherence and Logic"	96
16. Frequency of Grades for the Last 10 Experts When They Graded Criterion Number Two--"Development of Ideas".	97
17. Frequency of Grades for the Last 10 Experts When They Graded Criterion Number Three--"Diction"	98
18. Frequency of Grades for the Last 10 Experts When They Graded Criterion Number Four--"Emphasis"	99
19. Frequency of Grades for the Last 10 Experts When They Graded Criterion Number Five--"Organization Through Sentence Structure and Paragraphing"	100
20. Frequency of Grades for the Total Markings of the Five Criteria by the First 20 High School English Experts	101
21. Calculation of the Coefficient of Concordance, the Data Consisting of the Ranking of the Five Criteria by Montana High School English Experts.	103
22. Calculation of the Coefficient of Concordance, the Data Consisting of the Ranking of the Five Criteria by Nine College English Experts	104

VIII

ABSTRACT

The belief that valid and reliable criteria for the evaluation of English composition at the twelfth grade level could be established led to this investigation. An examination of literature revealed that during the first two decades of the twentieth century emphasis was placed upon comparison scales. In the attempts to establish criteria after 1920, numerous criteria were presented by different studies. In the studies where the criteria were tested for reliability and validity, the correlation coefficients were found to be low.

Criteria in this study were selected from nine sources in literature which were published from 1951 to 1962 by The National Council of Teachers of English, state English associations, national testing services, and individuals of national importance in the field of English. The criteria were ranked in terms of importance by nine college and 22 high school experts. Ranking was tested at the one per cent level for validity by Kendall's coefficient of concordance, "W," and agreement was found. The ranking was as follows: (1) Development of Ideas, (2) Coherence and Logic, (3) Organization Through Sentence Structure and Paragraphing, (4) Diction, and (5) Emphasis. To determine reliability of the criteria, 30 high school English experts from 26 different high schools rated, according to each criterion, 20 themes which were selected from 256 written by the seniors of four Montana high schools. Fourteen tests were made to test at the five per cent level the ratings given each theme according to each criterion by each of the 30 experts by using an analysis of variance test and chi square tests. A chi square test was also used to test at the five per cent level the overall grades given the themes. The homogeneity of the mean grades that teachers give each theme, of the grades that students received on themes, and of the total markings given by the experts was tested. Each criterion was tested separately for a group of 20 experts and a second group of 10 experts whose ratings were received about a month after the first ratings. Agreement was found for "Organization Through Sentence Structure and Paragraphing" for the group of 20 experts and for "Emphasis" for the group of 10 experts. Since this was the only agreement, reliability of the criteria for the evaluation of English composition was not established.

Validity of the criteria for the evaluation of twelfth grade English composition was established by agreement in the ranking in terms of importance by college and high school English experts.

Perhaps the only conclusion justified by this study is that, although teachers of English composition may feel that criteria are important in evaluating themes, there is no evidence of consistency in the employment of such criteria.

CHAPTER I

INTRODUCTION

It may be argued that a student has difficulty in becoming proficient in his writing when there is disagreement among his instructors in the way they rate written composition. A student may have one English instructor with one set of standards for evaluation, but his next English teacher may use a completely different means for judging compositions. Students are often expected to progress and to develop in their writing with different standards used by different teachers.

Two eminent educators with divergent views, John Dewey¹ and Robert M. Hutchins,² agree that teachers have a duty to young people that goes far beyond the traditional conception of the teacher's place in the school. They contend that before acceptance, appreciation, and learning can be expected, rapport must be established between the student and his teacher. In the relations between the English teacher and his students in written composition, it is very essential that this rapport exist if improvement in the ability to write is to occur. The belief long held by the writer has been that the availability of suitable criteria for the evaluation of English composition would bring about a type of standardization in the evaluation of written composition that would develop this rapport and which would result in improvement of students' ability to write. The

¹Dewey, John, The School and Society.

²Hutchins, R. M., No Friendly Voice.

desire to find these suitable criteria was the motivating force in this investigation.

Statement of the Problem

The major purpose of this study was to develop a set of criteria for the evaluation of English composition. In the development of the criteria, there were three major considerations: (1) selection of criteria, (2) establishment of validity, and (3) establishment of reliability. Once the validity and reliability of the criteria were established, the criteria could be considered as a useable instrument or scale which would give a standard means for the evaluation of English composition.

Procedures

The procedures used to select criteria, establish validity, and establish reliability are treated in detail in the following sections.

Selection of criteria. Selection of criteria was made from nine sources³ which were carefully selected from literature. The criteria

³The sources from which the criteria for the scale were devised are as follows: (1) California Association of Teachers of English, A Scale for Evaluation of High School Student Essays; (2) College Entrance Examination Board, "Report of the 1961 Western Advanced Placement English Conference," Occidental College, Los Angeles, California; (3) Dusel, W. J., "How Should Student Writing be Judged?" English Journal 46:263-8, May, 1957; (4) Educational Testing Service, A Brief--Cooperative.

these nine sources contained were considered representative of criteria for evaluating compositions. The criteria were then sent to college English experts throughout the United States to be ranked.

Establishment of validity. The selected criteria were ranked by the college experts, and their numerical rankings were treated statistically. In order to establish validity of the criteria, Kendall's coefficient of concordance, "W", was used to find the amount of agreement among the experts in the ranking of the criteria. High School English teachers were also selected to rank the criteria. Their numerical rankings were tested by Kendall's coefficient of concordance, "W".

Establishment of reliability. Thirty teachers were selected to grade themes by marking each theme (1) poor, (2) average, and (3) better than average according to each criterion, and the consistency with which they graded was determined by statistical analyses of their numerical grades. To determine reliability an analysis of variance test was used to test the differences in the grading of high school English teachers, and the chi square test was used to determine the significance of the grading by high school English teachers.

Sequential Tests of Educational Progress; (5) Grose, L. M., Miller, Dorothy, and Steinberg, E. R., editors, Suggestions for Evaluating Junior High School Writing; (6) Illinois Association of Teachers of English, Illinois English Bulletin, "Evaluating Ninth-Grade Themes," and "Evaluating Twelfth-Grade Themes;" (7) Indiana Council of Teachers of English, "Standards for Written English in Grade 12," Indiana English Leaflet, 3: 1-24, October, 1960; (8) National Council of Teachers of English, "Ideafom Paper;" and (9) University of the State of New York, The State Education Department, Suggestions on the Rating of Regent's Examinations in English.

Themes which were used in the study were obtained from four Montana high schools. The subject, "Neighbors, Good and Bad--People," was one for which the writer believed that the students had equal preparation. Length of the themes and time allowed for writing were set by the writer in order to prevent the themes from being either too long or too short and to give each student equal opportunity to handle his subject. Grade twelve was chosen to write the themes because twelfth grade students were considered to have a great interest in the people about them. Care was also taken to obtain an equivalent number of themes written by members of each sex. Every thirteenth theme was selected to be graded according to the specified criteria. The themes were not sorted in any particular fashion before selection but were mimeographed so that prejudices due to any extraneous factors would be eliminated.

Teachers were chosen from Montana high school English staffs to judge the compositions. The mimeographed themes, as well as a copy of the prepared criteria, were given to the high school teachers for evaluation. These teachers were asked to numerically rate the compositions according to the criteria and to make any comments pertinent to their evaluation. To determine reliability statistical analyses were used on the numerical grading of the themes by high school English teachers.

Limitations and Controls

In this study, there were certain limitations. The writing of

themes was limited to twelfth grade students from four high schools because at the twelfth grade level high school students are considered to be at their highest level of development, and a sufficient number of them themes for this study was obtained from this group. No attempt was made to control the way teachers conducted their classes aside from the specific instructions for writing themes.

Controls placed on the study were (1) that each of the high school teachers possess a Bachelor of Arts, Bachelor of Science, or higher degree, and have a major in English; (2) that each teacher have at least two years' teaching experience in the field of English; (3) that the teachers prepare students to write the themes according to specific instructions; and (4) that the limit of 200-300 words be observed.

It was considered an important first step to show what literature had revealed regarding the general status of the evaluation of English composition and to select criteria suitable for the evaluation of English composition. The results of this survey are presented in Chapter 2.

CHAPTER II

REVIEW OF LITERATURE

Many attempts have been made to bring about uniformity and objectivity in the grading of English compositions. By 1903, composition had come into its own in the schools, and serious questions about dependability of theme grading and the factors that go to make up a good theme began to be considered. Early graders, while assuming that theme criticism aids the student, did little to unify their standards. Inconsistencies among judges or with one judge when evaluating at different periods of time were prevalent. The lack of reliability or validity in grading began to be studied at the beginning of the twentieth century. After 1903, serious efforts were made to establish consistency in theme grading.

Between 1903 and 1950, two general approaches were made in an attempt to establish a grading scale for English composition. One approach consisted of developing a scale by combining grades of several or many readers to establish standards. The scales developed in this manner were called comparison scales. The other approach used to establish standards was to arbitrarily set up standards and have them tested by graders.

After 1951, many English experts made efforts to make grading of English compositions more uniform. Some of these people wrote articles which indicated that they were content to show inconsistencies, while others attempted through demonstrations of theme grading to indicate the

proper manner of grading. Some authorities provided criteria for judging themes but did not establish reliability or validity for their criteria.

Scale development is historically separated into two major divisions because of the interruption of activity due to World War II and its aftermath. Renewed activity was not vigorously commenced until after 1951; therefore, the studies that took place in the late 1940's were placed with the earlier group. The review of literature from 1903 to 1950 is now presented.

Attempts at Developing a Scale for the Evaluation of Compositions, 1903-1950

In 1903, Rice¹ selected typical themes as a result of combining grades of several readers for one theme and used these themes as examples for judging other themes. He read a story to more than 8,300 pupils of various schools and asked the pupils to reproduce the story in writing. The themes were graded by a process of sorting into five groups and, from this, computing distributive class averages. Sample copies of the reproduced story were selected as guides for further scoring. These copies of the reproduced story were used as a comparison scale by Rice. He found that when teachers used his comparison scale agreement among graders was fairly high. In 1904, Rice, after "a long and circuitous

¹Rice, J. M., "The Results of a Test in Language," The Forum 3:269-273, October, 1903.

search for a scientific system of marking composition,² was convinced that a theme must be judged as a unit.

In 1911, Thorndike³ developed a comparison scale for the measurement of quality in handwriting. A year later Hillegas⁴ published a comparison scale and used methods similar to those of Thorndike to devise his scale. Instead of having many pupils write papers to be judged by a few judges as Rice had done, Hillegas had over 400 competent judges determine quality in compositions by comparing themes of variant quality. In determining his scale, Hillegas also devised some sample themes for grading. According to Trabue,⁵ the Hillegas scale was one of the most useful measuring instruments in the whole field of education. In 1917, Trabue⁶ devised a further modification of the Hillegas scale by using actual themes of school children instead of devising some samples as Hillegas had done.

Two scales which used uniform intervals between samples were

²Rice, J. M., "The Need of a New Basis in Education," The Forum 35:443, January, 1904.

³Thorndike, E. L., "A Scale for Merit in English Writing by Young People," Journal of Educational Psychology 2:361-368, September, 1911.

⁴Hillegas, M. B., "A Scale for the Measurement of Quality in English Composition by Young People," Teachers College Record 13:5-8, September, 1912.

⁵Trabue, M. R., "Supplementing the Hillegas Scale," Teachers College Record 18:53-55, January, 1917.

⁶Ibid., pp. 53-55.

published by Hudelson⁷ in 1923. Huxtable⁸ between 1929 and 1935 prepared a scale concentrated on thought content which she divided into five levels. Each level was divided into two or more kinds of thought content, and each kind "defined" in evaluative terms but without numerical dimensions. A specialized scale, which was similar to that of Hudelson's, was developed in 1934 by Stewart⁹ for rating high school journalism articles.

During 1935, Anderson and Traxler¹⁰ made an attempt to relate the components of writing to each other and to the whole. However, the assignment of values was purely arbitrary. They used a ten-point system wherein a weight of two was given to "completeness and accuracy," three to "mechanics," and five to "coherence." Using the same themes, they then developed a 60 point scale in which six points each were given to "accuracy," "completeness," "spelling," "punctuation," and "language errors," and 10 points each for "coherence between main divisions,"

⁷Hudelson, Earl, English Composition, Its Aims, Methods and Measurement, pp. 77-78.

⁸Huxtable, Z. L., "Criteria for Judging Thought Content in Written Composition," Journal of Educational Research 19:190, March, 1929.

⁹Stewart, Marietta, "A Scale for Measuring the Quality of Conventional News Stories in High School Journalism," English Journal 23:221-222, March, 1934.

¹⁰Anderson, H. A., and Traxler, A. B., "Reliability of an Essay Test in English." (As related in Anderson, H. A., and Traxler, A. B., "The Reliability of the Reading of an English Essay Test: A Second Study," School Review 48:523-525, September, 1935.).

"organization of paragraphs," and "organization of sentences." Only the items of "spelling" and "punctuation" even approached objectivity, and the remaining items depended, as would be expected, almost entirely upon the grader's judgments.

In the June, 1939, "Comprehensive Examination in English" at the University of Chicago, a scale was used. The total possible grade on any theme was eleven, with three points awarded to "organization and paragraph structure," four to "accuracy in writing and technique of composition," and four to "general knowledge and skill in the use of books required by the topic."¹¹

McKean¹² in 1947 wrote an article entitled "If the Shoe Fits" in which he asked for clear, meaningful sentences in which there was interrelation among both sentences and paragraphs. He said, "Some instructors, who must run the risk of being called fanatics, could even go on to demand that freshman writing avoid errors of fact."¹³ He referred, of course, to the writing of college freshmen, but their writing is closely related to that of high school seniors.

¹¹Noyes, E. S., "Recent Trends in the Comprehensive Examination in English," Educational Record (Supplement 13) 21:118, January, 1940.

¹²McKean, Keith, "If the Shoe Fits," College English 8:260, February, 1947.

¹³Ibid., p. 260.

Wiseman¹⁴ in an article written in 1949 provided evidence of the possibility of developing reliable marking of themes. He reported that when four experienced persons graded the same 200 papers twice and separated their grading by a time interval of four months there was a correlation of .946 between the aggregate marks.

Travers¹⁵ in 1948 stated that evaluation procedures were "still in their infancy," and that there was a pressing need for the development of new appraisal methods for composition. Huddleston¹⁶ bluntly said a few years later that there had so far appeared no evidence to indicate that adequate essay tests could be devised.

The early research (1903-1939) was filled with reports of attempts to secure increased validity of essay tests and improve reliability among judges, but none of them proved to be overwhelmingly satisfactory. A noticeable decrease in the number of investigations in theme grading took place during the years of World War II. During the 1950's and early 1960's; the number of efforts to find a suitable scale for grading quality of English composition greatly increased.

¹⁴Wiseman, Stephen, "Grammar School Selection: Marking an Essay Reliably," Times Educational Supplement 1771:213, April 9, 1949.

¹⁵Travers, R. M. W., "A Review of Procedures for the Evaluation of the Outcomes of Teaching English," Journal of Experimental Education 17: 331, March, 1954.

¹⁶Huddleston, E. M., "Measurement of Writing Ability at the College Entrance Level: Objective vs. Subjective Testing Techniques," Journal of Experimental Education 22:204, March, 1954.

Scale Developments, 1951-1962

The attempts to make grading of English composition uniform continued in the era, 1951-1962, with renewed emphasis. Huddleston¹⁷ in 1952 used a rating system which evaluated themes for "material and organization," "spelling," "punctuation," "syntax," and "sentence structure." During the same year, Coward¹⁸ reported on a similar rating scale used to judge essay examinations for foreign service applicants. It included the seven variables of "material," "organization and style," "spelling," "punctuation," "grammar," "diction, and "rhetoric,"

Two booklets were published by the Illinois Association of Teachers of English¹⁹ in 1953. While one was designed to help in the evaluation of ninth-grade themes and the other for judging twelfth-grade themes, they were both much alike in nearly all respects. The booklets stated a common belief that one of the responsibilities of an English teacher is to evaluate composition--"to weigh and find what is wanting, to attach a symbol that tells the student that his work is good or poor, and to write a comment that will help the student to improve his

¹⁷Ibid., p. 190.

¹⁸Coward, A. F., "A Comparison of Two Methods of Grading English Composition," Journal of Educational Research 46:82, October, 1952.

¹⁹Illinois Association of Teachers of English, Illinois English Bulletin, "Evaluating Ninth-Grade Themes," and "Evaluating Twelfth-Grade Themes."

future writing in some way."²⁰ "If students are to become better writers, they must write, and what they write must be judged by a sympathetic and competent critic who knows what good writing is and who knows what students' capabilities are."²¹ The Illinois booklets contained checklists and the graders of themes were asked to please check "good," "fair," or "poor" for every item. The eleven criteria were:

1. Quality of content
2. Originality of treatment
3. Unity
4. Coherence
5. Emphasis
6. Paragraphing
7. Diction
8. Grammatical usage
9. Sentence structure
10. Spelling
11. Punctuation

Sample themes with markings, comments to the student, final grades, and comments for teachers appeared in the pamphlets, and the committee believed that this material could be used both by teachers and by students.

²⁰ibid., p. 2.

²¹ibid., p. 2.

In an Educational Testing Service Bulletin of 1955, Diederich²² related that few men of genius have excelled in criticism and that their judgments were not infallible in regard to certain words. Diederich listed and explained what he referred to as artificial difficulties that graders created for themselves. "Cocksureness," he defined as the process of beginning with modesty and fear, but after grading 20 papers starting to feel quite sure that no other judgment except one's own was correct. He went on to relate: "The practice fills the judge with sinful pride. When we snort violently and say, 'This, by God is an F,' we should be particularly on guard. Perhaps God is not a partner to the decision."²³ Other downfalls according to Diederich are "timidity," "hypersensitivity to any certain fault or good in a paper," and the "question-answering concept of an examination."²⁴ However, he felt that it was considerably more difficult to state things to do. A letter grade for composition was determined by Diederich by sorting the papers into five groups with concentration on A's, C's, and F's. B's and D's were considered borderline cases. He attempted to explain what a paper should contain at each of the three main letter grading levels. Although he established no scale and claimed throughout that his methods of grading were

²²Diederich, P. B., Notes on Grading Essays, Educational Testing Service, November 22, 1955, pp. 1-6.

²³Ibid., p. 1.

²⁴Ibid., pp. 2-3.

devised from his personal opinion; Diederich did give information and instruction to the average grader.

During that same year, 1955, Wolfe, Geyer and others sent to press the second edition of an English textbook, Enjoying English. Within the bounds of this text were several evaluation charts. The authors point out that the student should check his topic sentences, paragraphs, variation of sentences, use of dialogue, periods, and spelling. The following "Self-Judgment Chart" was designed for the topic, "Friendship." It was not only an example of all the charts but was pertinent to the topic as a whole.

1. In your topic sentence have you given your attitude toward the friendship you are discussing?
2. Are you specific in your paragraph? Have you told the important facts: where you met your friend, who he is, the details that explain your attitude in the topic sentence?
3. Are your sentences varied in pattern? Do some open with phrases or clauses telling when or where? In at least one sentence, after an opening phrase have you used a dependent clause beginning with when, while, or as?
4. Have you used dialog where it helps to characterize your friendship?
5. Have you spelled correctly such words as description, acquaintance, existence, occurrence, dissappoint, possession, experience, occasion, loneliness, separate, humorous?²⁵

The book was designed to give various "Self-Judgment Charts" for different levels.

Also during 1955 the National Council of Teachers of English

²⁵Wolfe, D. M., Geyer, E. M., and others, Enjoying English, p. 105.

produced what was known as "Ideaform Paper." On one side of this paper were lines where the theme was to be written. On the other side, there was space provided for an overall grade, comments on the ideas or content of the theme, other comments, and a ranking of "good," "fair," or "poor" in terms of the evaluation criteria which were:

1. Organization
2. Development
3. Sincerity of purpose and expression
4. Sentence structure and punctuation
5. Paragraphing
6. Usage
7. Choice of words
8. Spelling
9. Penmanship and general appearance

"Ideaform Paper" was approved by both the High School Section Committee and the Immediate past chairman of the Conference on College Composition and Communication and was designed to give the student greater skill in the writing of English composition.

One other booklet came into being at approximately the same time when Wolfe, Diederich, The National Council of Teachers of English, and the Illinois committee were making their various studies. This booklet

which was titled Suggestions for Evaluating Junior High School Writing²⁶ was organized to provide a frame of reference to which a beginning teacher could refer as a basis for grading English composition and to provide the experienced teacher with sample themes which were corrected and graded by other teachers so that she would know how her grading compared with theirs. The booklet contained themes for seventh, eighth, and ninth grade levels, but no attempt was made to regulate the approach or bring about exacting consistency. Three main principles for evaluating junior high school themes were enumerated in the booklet.²⁷ These three main principles were content, form, and mechanics under each of which appeared the following questions:

CONTENT

1. Does the student have something to say?
2. Has the student expressed his ideas with sincerity and conviction and an earnest desire to communicate?
3. Does the student display a touch of freshness and originality in his viewpoint, offer an idea, an observation that is distinctly his own?
4. Are there traces of the student etched into the theme that give it substance and vitality?

FORM

1. Does the theme have unity?

²⁶Grose, L. M., Miller, Dorothy, and Steinberg, E. R., editors, Suggestions for Evaluating Junior High School Writing.

²⁷Ibid., pp. 4-5.

2. Does the theme have coherence?
3. Is the theme effective?

MECHANICS

1. Is there freedom from fragmentary or run-together sentences, and from breakdown or muddled sentences?
2. Is there freedom from dangling or misplaced modifiers; from constructions showing lack of agreement of subject and verbs, of noun or pronoun and antecedent; from mutilated verb forms; and from inconsistencies such as shift in person, mood, voice, or tense?

In 1955, Thomas²⁸ published a booklet designed to aid teachers in correcting themes. She used many samples of corrected themes as a basis for future evaluation. No evaluation scale or set of specific criteria was established.

In a Kentucky English bulletin²⁹ published in 1956, no criteria as such were established but the author's opinion on the proper method of correcting themes was demonstrated by showing many theme examples that contained grading notes and comments. In the preface was stated what was expected of the writer of a theme. The relative values which must be considered by the grader and methods for the reader to use were also given. These values consisted of "contents vs. mechanics," whether a "system of penalties" was sufficient, "pros and cons in the use of

²⁸Thomas, E. S., Evaluating Student Themes.

²⁹Ward, W. S., editor, "Principles and Standards in Composition for Kentucky High Schools and Colleges," Kentucky English Bulletin.

marginal notations," and "the use of terminal comments." It was concluded that, while the terminal comment was probably of most use to the student, the grader of compositions should also make marginal comments and mark mechanical errors.

The Educational Testing Service in 1957 published a bulletin³⁰ containing an English composition evaluation scale. Under the sub-title, "What are the Writing Skills Tested?" were listed the following:

1. The ability to express oneself logically in writing.
2. The ability to organize materials: in the whole passage, in the paragraph, and in the sentence.
3. The ability to write appropriately: to the purpose, the occasion, the audience.
4. The ability to write effectively: word choice, emphasis, details, reasons, examples, comparisons, exactness and clarity, simplicity, economy, variety, imagination, force. (This list is not exhaustive nor does it mean that all characteristics should be expected in one selection.)
5. The ability to use the conventions of writing: in basic form mechanics, in syntax mechanics, in word-form mechanics.

It was also noted in this bulletin of the Educational Testing Service that emphasis should be placed on the total writing process and not be given to minor matters, gross errors, or to the niceties of usage.

Dusel,³¹ in an article published in May, 1957, suggested four

³⁰ Educational Testing Service, A Brief--Cooperative Sequential Tests of Educational Progress, p. 12.

³¹ Dusel, W. J., "How Should Student Writing be Judged?" English Journal 46:263-8, May, 1957.

standards of judgment for written expression. These were as follows:

1. Honesty and sincerity
2. A growing awareness of complexity
3. Order in the pupils' writing
 - a. Appropriateness in style
 - b. Correctness in style--diction that deals with boys and girls, sticks and stones.

Dusel also insisted on standard spelling, standard punctuation, and standard sentence patterns for written expression. He stated that we should respect accuracy in writing--"The ability to reveal to the reader exactly what the writer intends to reveal and no more."³² Dusel seemed to be keenly aware of the complexity of grading as shown in his statement: "If only we could judge writing without having to face the writer the next day--or having to teach them to write better--how easy the job would be!"³³ Other than suggesting and explaining the four points, he more or less avoided any type of solid-base for evaluation. The article seemed to indicate that Dusel's purpose was to force realization of grading variation and offer guidance, but leave the final decisions up to the individual.

In 1958, Miller developed, used, and tested grading scales. The two-fold purpose was to develop a reliable method of grading freshman

³²ibid., p. 265.

³³ibid., p. 264.

themes and, at the same time, to determine the improvement in writing made by freshman English students at the University of Minnesota. Themes were given an over-all or composite grade. Grading was also done according to the scales Miller developed. After grading took place, it was found that graders disagreed significantly on all but the comma splice and dangling modifiers, as shown in Miller's statement:

A rating scale was devised, including an over-all grade, three items of structure, three of content and presentation, and seventeen errors in mechanics, or weaknesses in construction. The scale was applied by five graders to 100 randomly selected impromptu themes written in 1949-1950 and 1950-1951. The median correlation between graders for the over-all grade was .375, and the reliability coefficient .73. The variability among the graders was even greater for the other items.

The scale was revised to contain an over-all grade, six items of quality of writing, and seven mechanical errors, and applied independently by three graders to 200 randomly selected pairs of themes, written at the beginning and end of the school year. The median correlation for grades and qualities was .46, and the median reliability coefficient .73. There were statistically significant disagreements among the graders on all but two of the items, the comma splice and dangling modifiers.³⁴

Miller concluded that neither of the two rating scales he devised achieved satisfactory reliability. He stated that: "The graders differed widely in their evaluations of the over-all grade and the qualities of writing, and those errors which could be the least frequent and the

³⁴Miller, J. W., An Analysis of Freshman Writing at the Beginning and End of a Year's Work in Composition, Doctoral dissertation.

least important in affecting writing quality."³⁵

A series of conferences held throughout 1958 by members of the American Studies Association, College English Association, Modern Language Association, and National Council of Teachers of English brought into being a booklet³⁶ designed primarily to define and clarify the possibilities, limitations, and desired effects on methods of grading English compositions. No effort was made in the booklet to answer the question of whether or not national standards for students' writing in various grades could be established or what their value would be; however, the problem of the grader and the consequent misconceptions that arise in the minds of pupils were presented. The problem of the grader, as stated in the following quotation, was whether to mark mechanical and grammatical errors only, to mark very subjectively, or to use a combination of error marking and subjectivity:

... Some overworked teachers mark only mechanical and grammatical errors, leaving the students with the impression that learning to write well is a negative matter--the avoidance of such errors. Others go too far in the other direction and grade very subjectively, leaving the student with the impression that the art of writing well is merely the knack of appealing to the tastes and whims of his particular teacher.³⁷

³⁵Ibid., p. vii.

³⁶American Studies Association, and others, The Basic Issues in the Teaching of English. (Published as a supplement to College English 21:1-15, October, 1959.)

³⁷Ibid., pp. 9-10.

Vickery³⁸ undertook a study in 1959 to establish as objective a scoring system as possible to be used as part of a composition testing device in the North Dakota State Teachers College's testing program. He³⁹ advocated a liberal system based upon the "Willing Composition Scale." He stated that this scale provided for evaluation of compositions for "story value" and "form value:"

Story value is determined by comparing the papers to be evaluated with a series of eight samples of student writing of increasing value, numbered 20, 30, 40, 50, 60, 70, 80, and 90. Value indexes are assigned according to the number of the sample which a given paper most closely resembles. Form value is determined by noting all errors in usage in a given paper. The total number of errors is multiplied by 100 and divided by the number of words. The resulting quotient indicates the number of errors per 100 words.⁴⁰

Vickery decided to disregard "story value" and to concentrate on "form value" or technical competence. At the time his introduction was written the program had been in operation for five quarters and seemed to be proving successful.

Christ⁴¹ recommended the materials published by the National

³⁸Personal letter written by Dr. Kenton F. Vickery, North Dakota State Teachers College, Dickinson, North Dakota, to the writer on March 15, 1962."

³⁹Vickery, K. F., The Use of Objective Measurement of Writing Techniques, (Published in the North Dakota Teacher, September, 1959).

⁴⁰Ibid., p. 3.

⁴¹Personal letter written by Henry I. Christ, Head, English Department, Andrew Jackson High School, New York, N. Y., to the writer on May 15, 1962."

Council of Teachers of English in reply to a request for information about the grading of English composition. In 1960, Tressler and Christ were co-authors of an English text, English in Action, which was published with variations on the ninth, tenth, eleventh, and twelfth grade levels. This series contained a wealth of tables and scales designed to improve student writing. In checking for good paragraphs, the authors pointed out that the paragraphs should be worth reading, that the paragraphs should be clear, that they should have a strong beginning and end, and that the paragraphs should hang together. Spelling and grammar should be correct. The following questions which appear in Chapter 5, English in Action, Course Two, are called a "Check List for Good Paragraphs:"

1. Have you something to say that is worth reading?
2. Is the point you are trying to make perfectly clear? Is your topic narrow enough?
3. Will your opening arouse curiosity, interest, agreement, or even antagonism?
 Weak beginning: I don't like fashions, and I wish I could change them.
 Better beginning: Last night I dreamed I was a fashion dictator.
4. Do you keep to the point?
5. Have you used pointed details, forceful examples, striking comparisons, and convincing reasons?
6. Does the paragraph hang together? Are the details arranged in logical order?
7. Is your ending forceful, or does it sputter to a stop?
 Weak ending: Those are some of the reasons why I like hockey.
 Better ending: That's hockey--the world's fastest sport and my favorite.
8. Is every word spelled correctly?
9. Is your grammar correct throughout?

10. Have you crossed out every unnecessary word?⁴²

Examples were used by Tressler and Christ to clarify for the student the meaning of certain criteria.

In 1960, a study⁴³ involving the objective analysis of composition was made which was concerned with students' writing when they were writing about literature. The author suggested that the instructor begin by considering the student's presentation in regard to "process," "synthesis," and "originality." The first of these "process," Ruoff indicated as the problem of whether or not the student's argument was substantiated and fully developed when he said:

... whether the student's argument--and every literary analysis is, ultimately, an argument--is adequately substantiated by textual reference and allusion, fully developed as an argument, logically deduced as a series of logical propositions, and, perhaps most important, expressly cognizant of all the complexities pertinent to the problem treated in the essay?⁴⁴

Ruoff related that synthesis demanded that the student "...take into account the salient features of his problem and bring these together into a significant relationship."⁴⁵ He indicated that knowledge, together with

⁴²Tressler, J. C., and Christ, H. I., English in Action, Course Two.

⁴³Ruoff, James, "Evaluating Student Essays in Literature: A Plea for Objective Criteria," College English 22:35-7, October, 1960.

⁴⁴Ibid., p. 36.

⁴⁵Ibid., pp. 36-37.

discovery and revelation, made a composition original. The following statement brings out the importance he attached to discovery:

Finally, consistent with the principle of originality the student is required to propose in his essays answers suggestive of something beyond, and perhaps even different from the instructor's expressed formula or Cleanth Brooks's ingenious dissection, and this on the sound theory that literature, since it entails not facts but values, exists above the pedestrian level of mere information, and ought therefore to pertain to knowledge, which has to do with discovery and revelation--with what is experiential and uniquely personal.⁴⁶

Ruoff's criteria in many ways were found to agree with the ideas of Dusel.⁴⁷ While the criteria were not exactly the same, either set could be of importance to the student in writing compositions.

Appearing in 1960, was a pamphlet⁴⁸ prepared by the California Association of Teachers of English which stressed bringing increased coherence to the English teacher's job of improving student writing. In this pamphlet were listed topics and sub-topics, and the pamphlet carefully explained questions that should be defined so that they might be used in the evaluation of essays. The three main criteria appearing in the pamphlet of the California Association of Teachers of English were "content," "organization," and "style and mechanics." The three were

⁴⁶Ibid., p. 37.

⁴⁷See Chapter 2, pages 17 and 18 for the ideas of Dusel.

⁴⁸California Association of Teachers of English, A Scale for Evaluation of High School Student Essays.

presented as follows:

- I. Content: Is the conception clear, accurate, and complete?
 - A. Does the student discuss the subject intelligently?
 1. Does he seem to have an adequate knowledge of his subject?
 2. Does he avoid errors in logic?
 - B. Does the essay offer evidence in support of generalization?
- II. Organization: Is the method of presentation clear, effective, and interesting?
 - A. Is it possible to state clearly the central idea of the essay?
 - B. Is the central idea of the paper as a whole sufficiently developed through the use of details and examples?
 - C. Are the individual paragraphs sufficiently developed?
 - D. Are all the ideas of the essay relevant?
 - E. Are the ideas developed in logical order?
 1. Are the paragraphs placed in natural and logical sequence within the whole?
 2. Are the sentences placed in natural and logical sequence within the paragraphs?
 - F. Are the transitions adequate?
 - G. Are ideas given the emphasis required by their importance?
 - H. Is the point of view consistent and appropriate?
- III. Style and Mechanics: Does the essay observe standards of style and mechanics generally accepted by educated writers?
 - A. Are the sentences clear, idiomatic, and grammatically correct? (For example, are they reasonably free of fragments, run-on sentences, comma splices, faulty parallel structure, mixed constructions, dangling modifiers, and errors of agreement, case, and verb forms?)
 - B. Is the sentence structure effective?
 1. Is there appropriate variety in sentence structure?
 2. Are uses of subordination and coordination appropriate?
 - C. Is conventional punctuation followed?
 - D. Is the spelling generally correct?

E. Is the vocabulary accurate, judicious, and sufficiently varied?⁴⁹

The California Association also formulated symbols to be used in marking the essays and demonstrated with the use of several compositions the principles of both the criteria and the markings.

Another pamphlet⁵⁰ of importance was published in Indiana during 1960. The main purposes of the authors of this pamphlet was to suggest criteria for the grading of written work. The scale presented in the pamphlet was broken into "content," "organization of the whole theme," "sentence structure," "diction," and "grammar, punctuation, spelling." A table was presented in the pamphlet in which the superior, average, and unacceptable categories were described as follows:

CONTENT

Superior (A-B)	A significant central idea perhaps imaginatively conceived, but at any rate clearly defined, and supported with concrete, substantial, and consistently relevant detail
Average (C)	Central idea apparent but trivial, or trite, or too general; supported with concrete detail, but detail that is occasionally repetitious, irrelevant, or sketchy
Unacceptable (D-F)	Central idea lacking, or confused, or unsupported with concrete and relevant detail

⁴⁹Ibid., p. 13.

⁵⁰Indiana Council of Teachers of English, "Standards for Written English in Grade 12," Indiana English Leaflet, 3:1-24, October, 1960.

ORGANIZATION OF THE WHOLE THEME

Superior (A-B)	Theme planned logically, so that it progresses by clearly ordered and necessary stages; and developed with originality and consistent attention to proportion and emphasis; paragraphs unified and effectively developed; transitions between paragraphs explicit and effective
Average (C)	Plan and method of theme apparent but not consistently fulfilled; developed with only occasional disproportion or inappropriate emphasis; paragraphs unified and generally effective in their development; transitions between paragraphs weak or mechanical
Unacceptable (D-F)	Plan and purpose of theme not apparent; undeveloped or developed with irrelevance, redundancy, or inconsistency; paragraphs incoherent, not unified, or undeveloped; transitions lacking

SENTENCE STRUCTURE

Superior (A-B)	Sentences skillfully constructed (unified, coherent, forceful, effectively varied)
Average (C)	Sentences correctly constructed but lacking distinction
Unacceptable (D-F)	Sentences not unified, fused, incomplete, monotonous, or childish

DICTION

Superior (A-B)	Distinctive: fresh, precise, economical, and idiomatic
Average (C)	Appropriate: clear and idiomatic
Unacceptable (D-F)	Inappropriate: vague, unidiomatic, or substandard

GRAMMAR, PUNCTUATION, SPELLING

Superior (A-B)	Clarity and effectiveness of expression promoted by consistent use of standard grammar, punctuation, and spelling
Average (C)	Clarity and effectiveness of expression weakened by occasional deviations from standard grammar, punctuation, and spelling
Unacceptable (D-F)	Communication obscured by frequent deviations from standard grammar, punctuation, and spelling ⁵¹

As in other studies, the Indiana pamphlet gave examples in which use of these criteria was made.

An abundance of reports of work in the area of establishing criteria for correction of quality in English composition appeared during 1961. Lynch,⁵² in one report, related that a grade was based on content and style, as well as correctness. He felt that, in these areas--content and style--there was much indefiniteness in grading. Part of the difficulty in precision in the grading of content was attributed to the tendency of the elements of style and content to merge into each other. Lynch believed that the only criterion for evaluating content alone was the judgment of the teacher. He revealed that in the area of style such specific measures as these given in the following statement may be established:

⁵¹Ibid., p. 3.

⁵²Lynch, J. J., "The Achievement of Excellence in Composition," The High School Journal 45:16, October, 1961.

Originality, novelty, imagery, diction, and rhetorical devices, such as figures of speech, are then understood to be part of the evaluation and must be considered by the student for the exhibition of genuine excellence in writing.⁵³

In April, 1961, Koclanes,⁵⁴ after ample review of several notable studies such as that of Traxler and Anderson,⁵⁵ decided that it was possible to evaluate compositions if the meaning of evaluation were limited. Koclanes indicated by the following statement his belief that criticism could be given to help improve written composition if it is directed at improvement of ability and not grading:

If by evaluation, then, we mean criticism with the intention of teaching individuals how to improve their written expression, and not the grading of compositions, I believe that the question in the title of this paper ("Can We Evaluate Compositions?") can be answered in the affirmative.⁵⁶

Koclanes⁵⁷ apparently favored stress on ability over grading because he found that in essay testing the usual correlation of the grading of several teachers was low.

The College Entrance Examination Board's General Composition

⁵³Ibid., p. 16.

⁵⁴Koclanes, T. A., "Can We Evaluate Compositions?" The English Journal 50:250-257, April, 1961.

⁵⁵Trexler and Anderson, op. cit., pp. 534-539.

⁵⁶Koclanes, op. cit., p. 253.

⁵⁷Ibid., p. 252. (Citing Stalnaker, J. M., and Stalnaker, R. C., "Reliable Reading of Essay Tests," School Review 42:599-605, 1934.)

Test⁵⁸ for 1961 provided for a four-part rating scale for themes. The points on this scale were as follows:

1. Appropriateness of topic, not necessarily scientific but definitely not trivial
2. Logical development which went beyond mere neat paragraphing or obvious step by step development of process
3. Organizational pattern, such as definition, description of process, and consideration of result
4. Correctness and grace of style

The points on the rating scale were part of the detailed analysis of the standards by which the four examination questions that were given were scored and formed the basis for the defense against the only two major criticisms that had been made of the 1961 examination. At the 1961 Western Advanced Placement English conference, the two major criticisms were that one grade was given for literature and composition and that the prose analysis should not have been added. Borden⁵⁹ felt that the objection of a composite grade could be refuted by the number of students who handled the amount of materials well and by the need for a question on prose analysis. The examinations, therefore, stood firmly without

⁵⁸College Entrance Examination Board, "Report of the 1961 Western Advanced Placement English Conference," Occidental College, Los Angeles, California, p. 3.

⁵⁹These were comments of Professor Arthur R. Borden, Washington and Lee University, Lexington, Virginia, at the 1961, Western Advanced Placement English Conference, Occidental College, Los Angeles, California, on June 22-24, 1961.

changes being made.

Estrin⁶⁰ in December, 1961, wrote an article titled "How Do You Grade A Composition?" To answer the question in his title, he wrote the Departments of English in 100 colleges and universities. From the responses of 80 of these colleges he was able to develop a scale for correcting themes. In remarking about the scale, it was pointed out that gross errors such as comma splices and fused sentences will almost automatically cause a paper to fail. In the scale he listed the following criteria in descending order of importance:

1. Sentence structure
2. Paragraphing and paragraph development
3. Logical organization and outlining
4. Content
5. Spelling
6. Punctuation
7. Capitalization
8. Diction
9. Idiom
10. Originality and vigor

⁶⁰Estrin, H. A., "How Do You Grade A Composition?" College Composition and Communication 12:234, December, 1961.

The University of the State of New York supplied a pamphlet⁶¹ to New York high schools for rating the Regent's examination in English composition. This pamphlet included the criteria for evaluation of English compositions listed below:

1. Introduction and conclusion
2. Emphasis
3. Thought coherence
4. Word coherence
5. Design
6. Use of idioms
7. Word order
8. Vocabulary
9. Spelling
10. Mechanics

A grading form included per cent grading scheme from "excellent" to "rejected." Examples of theme grading were included in the pamphlet to give the grader of themes some idea on how to use the scale.

The television program, "College Bowl,"⁶² sponsored a theme

⁶¹University of the State of New York, The State Education Department, Suggestions on the Rating of Regent's Examinations in English, n.d.

⁶²"College Bowl," "Why I Want to go to College," (General Electric Series), produced by ABC Television Network, Broadcast by KXLF T. V., Butte, Montana, 3:30 p.m., February 25, 1962. Master of ceremonies for the series was Allen Ludden.

writing contest in 1962 which was open to all students in grades nine through twelve who would not be 21 before the closing date of the contest. The subject to be written about was "Why I Want to go to College," and the final judgment of the themes was made under the supervision of the editorial staff of World Book Encyclopedia. In the contest, 40 per cent of the grading was based on "organization and presentation of thought;" Another 40 per cent went to "clarity and originality of expression," and 10 per cent went to each of the following criteria: "correctness of spelling and appropriateness of language" and "legibility and neatness;"

During this period, 1951-1962, comparison scales were occasionally evident, but the trend seemed to be to attempt to establish specific criteria for evaluation of English composition.

In this survey of literature from 1951-1962, the writer felt that there were nine publications which presented criteria for the evaluation of English composition which were representative of all publications. These publications were prepared by what the writer considered some of the best qualified experts in the field of English in the United States since they represented state English associations, national testing services, the National Council of Teachers of English, and a nationally known individual in the field of English. These publications were considered very important because the individuals and groups who prepared them were of high prestige in the field of English and because the scales were to be

used or were used in the grading of compositions of a large segment of our student population. Two very important examples of the importance of the criteria are the scales of the College Entrance Examination Board and that of The Educational Testing Service. These scales were used in determining the future careers of many college students. The nine publications were:

1. California Association of Teachers of English, A Scale for Evaluation of High School Student Essays.
2. College Entrance Examination Board, "Report of the 1961 Western Advanced Placement English Conference," Occidental College, Los Angeles, California.
3. Dusel, W. J., "How Should Student Writing be Judged?" English Journal 46:263-8, May, 1957.
4. Educational Testing Service, A Brief--Cooperative Sequential Tests of Educational Progress.
5. Grose, L. M., Miller, Dorothy, and Steinberg, E. R., editors, Suggestions for Evaluating Junior High School Writing.
6. Illinois Association of Teachers of English, Illinois English Bulletin, "Evaluating Ninth-Grade Themes," and "Evaluating Twelfth-Grade Themes."
7. Indiana Council of Teachers of English, "Standards for Written English in Grade 12," Indiana English Leaflet, 3:1-24, October, 1960.
8. National Council of Teachers of English, "Ideaform Paper."
9. University of the State of New York, The State Education Department, Suggestions on the Rating of Regent's Examinations in English.

Summary

Composition scale development in the early twentieth century was

not considered successful since no scale became standard for judging written composition. Comparison scales were used to arrive at objectivity in grading but were very limited in their use and success. A comparison scale for measuring quality in English composition was established in 1912, and modifications of this scale were made in 1917. In 1923, two scales were published which used uniform intervals among samples as criteria. During the period, 1929-1935, in the development of criteria for judging quality of compositions, different types of thought content were found to be used as criteria. Thought content was broken down into a number of components in most of the studies after 1935, and work was done to establish reliability and validity for some of the scales without too much success. Scale items such as organization and paragraph structure, as well as accuracy and technique, were considered in such tests as the "Comprehensive Examination in English" given at the University of Chicago in 1939. During the years of World War II, little activity took place. After the war there was a noticeable rejuvenation of attempts to establish scales for the evaluation of English composition.

Between 1951 and 1962, many concerted efforts were made to establish criteria for the evaluation of English composition for the specific use of a particular group or for a particular test. Many scales for the evaluation of English composition were developed by groups and individuals but when the scales were tested, reliability and validity were found to be low. Since this was true, the writer decided to select criteria that

appeared in several competent sources in literature and test that criteria for reliability and validity with the expectation that the scale builders' concerted opinions on the importance of scale points were correct. It was felt that criteria from nine selective sources would be representative of all criteria in literature.

The basic difficulties in using any scale in the judging of writing has probably been the individual tastes of the graders and their disagreement as to whether a certain theme does or does not contain a certain criterion.

There has been a vast amount of literature written about the subject of whether or not criteria can or should be established for judging value in compositions, and, if so, what the various criteria should be. No two writers have actually reached agreement on the points to be included, but many of them have approached agreement by including many of the same ideas.

The major purpose of the present study was to develop a set of criteria for the evaluation of English compositions. The purposes of Chapter 2 were to investigate literature, to report the history, and to present selected criteria for the evaluation of English composition. Establishment of validity for the criteria was a major consideration. The results of this second step are presented in Chapter 3.

CHAPTER III

SELECTION AND VALIDATION OF CRITERIA FOR THE EVALUATION OF ENGLISH COMPOSITION

The selection of criteria for the evaluation of English composition was made from literature written by people of high prestige in the field of English. As a basis for this selection, nine sources¹ were chosen which contained criteria representative of criteria presented in literature.

Validation of the criteria was established by college English experts and high school English teachers who were experts in grading English composition. These experts ranked the criteria in order of importance and statistical analyses were made of their numerical rankings. If these experts agreed in the manner of ranking the criteria, then the criteria would be considered valid.

¹The sources from which the criteria for a scale were devised are as follows: (1) California Association of Teachers of English, A Scale for Evaluation of High School Student Essays; (2) College Entrance Examination Board, "Report of the 1961 Western Advanced Placement English Conference," Occidental College, Los Angeles, California; (3) Dusel, W. J., "How Should Student Writing be Judged?" English Journal 46:263-8, May, 1957; (4) Educational Testing Service, A Brief--Cooperative Sequential Tests of Educational Progress; (5) Grose, L. M., Miller, Dorothy, and Steinberg, E. R., editors, Suggestions for Evaluating Junior High School Writing; (6) Illinois Association of Teachers of English, Illinois English Bulletin, "Evaluating Ninth-Grade Themes," and "Evaluating Twelfth-Grade Themes;" (7) Indiana Council of Teachers of English, "Standards for Written English in Grade 12," Indiana English Leaflet, 3:1-24, October, 1960; (8) National Council of Teachers of English, "Ideafarm Paper;" and (9) University of the State of New York, The State Education Department, Suggestions on the Rating of Regent's Examinations in English.

The method of selection of the criteria was to survey all pertinent literature and select criteria. The method of establishing validity was to obtain agreement of experts in the ranking of criteria. The procedures involved in selecting and validating criteria are presented in the following sections.

Selection of Criteria for the Evaluation Scale

Criteria for the evaluation of English composition were selected from criteria presented by three state associations of teachers of English, from one state education department, from the National Council of Teachers of English, from two nation-wide testing services, and from one article written by a nationally known English expert. These were chosen because they all contained actual composition scales, were published and used in the ten year period just previous to this study, and contained criteria which were in English composition scales.

Although there were additional criteria presented in literature, a frequency count was kept while the survey was taking place, and the most frequently occurring items were selected. There were actually seven criteria selected, but because of similarity of meaning coherence and logic were combined into one criterion as was organization through sentence structure and paragraphing. The emphasis given each of the five selected

criteria in each of the nine sources² of national importance is shown in Table 1. Some of the selected criteria appeared as parts of other criteria in the nine sources. For example, the Indiana Council of Teachers of English placed emphasis under "Organization of the Whole Theme," and the California Association of Teachers of English placed logic under "Content: Is the conception clear, accurate, and complete?" and under "Organization: Is the method of presentation clear, effective, and interesting?" The five criteria which were chosen on the basis of frequency of occurrence in literature were as follows:

1. Coherence and Logic
2. Development of Ideas
3. Diction
4. Emphasis
5. Organization Through Sentence Structure and Paragraphing

None of the nine scales used as a basis for selection of criteria for the evaluation of English composition was tested for validity, but the writer felt that since the authors were of high prestige in the field of English that representative criteria selected from their suggested scale points could be tested and found to be valid. A detailed discussion of the selection of criteria follows:

"Coherence and Logic" includes thought and word coherence, logical

²See Chapter 3, page 39, for a list of the nine sources.

TABLE 1. EMPHASIS GIVEN FIVE CRITERIA IN THE NINE SOURCES

Source	Criteria as Listed in Sources				
	Coherence and Logic	Development of Ideas	Diction	Emphasis	Organization Through Sentence Structure and Paragraphing
Calif. Ass'n. of Teachers of English	Part of another criterion	Part of another criterion	Part of another criterion	Part of another criterion	Selected
College Entrance Exam. Bd.	Part of another criterion	Selected	Not selected	Not selected	Not selected
W. J. Dusel	Part of another criterion	Part of another criterion	Part of another criterion	Part of another criterion	Part of another criterion
Educational Testing Service	Selected	Part of another criterion	Part of another criterion	Part of another criterion	Selected criterion
Grose, Miller, and Steinberg	Selected	Part of another criterion	Not selected	Not selected	Part of another criterion
Ill. Ass'n. of Teachers of English	Selected	Selected	Selected	Selected	Selected
Ind. Council of Teachers of English	Part of another criterion	Part of another criterion	Selected	Part of another criterion	Selected
Nat. Council of Teachers of English	Not Selected	Selected	Selected	Not Selected	Selected
Univ., State of N.Y., State Educ. Dept.	Selected	Selected	Selected	Selected	Selected

planning of the whole theme, and intelligent discussion of the subject. Eight of the nine sources from which criteria were selected in this study either considered coherence and logic as one criterion or named them individually.³ Since coherence and logic were used by a majority of the experts in the literature considered, they were combined and selected as a criterion for the evaluation of English composition.

"Development of Ideas" was chosen as a criterion because it was selected as important by all nine of the sources, as shown in Table 1. This criterion includes overall effectiveness, content, introduction and conclusion, evidence in support of generalizations, originality of treatment, honesty in expression, and sincerity in expression.

"Diction" is used in this study to mean correct choice of words or appropriate wording. Because of its importance as expressed by the experts in seven of the nine sources, "Diction" was selected as one of the criteria for the evaluation of English composition.

"Emphasis" was considered important in six of the nine sources. The meaning of "Emphasis" as it was used as a criterion for the evaluation of themes can be best pointed out with the use of the question, "Has the student used the correct emphasis to make his writing appropriate to the purpose and the occasion?"

The criterion, "Organization Through Sentence Structure and

³See Chapter 2, pp. 6-38 and Table 1; page 42.

Paragraphing," was used in some form by all the sources except the College Entrance Examination Board. Sentence structure or paragraphing and in some cases both⁴ were used by the other eight sources. Since sentence structure and paragraphing are strongly related, these were incorporated into a single criterion.

The most frequently occurring criteria for the evaluation of English composition in the nine sources were (1) Coherence and Logic, (2) Development of Ideas, (3) Diction, (4) Emphasis, and (5) Organization Through Sentence Structure and Paragraphing. Experts were chosen from representative areas to rank the criteria. These experts were asked to rank the criteria in terms of its importance and statistical analyses of the numerical rankings were used to establish validity. Two groups, college English experts and high school English experts, were asked to rank the criteria in order of its importance.

Validation of Criteria by College English Experts

Validity of the five selected criteria for the evaluation of English composition would be established if there was agreement in their ranking by college experts. The methods used in the selection of experts, and the statistical treatment of their numerical rankings follow:

Selection of college experts. It was decided that the rankings of

⁴See Chapter 2, pp. 6-38 and Table 1, page 42.

approximately 10 college English experts would be necessary to test for validity of the criteria for the evaluation of English composition. Since statistical authorities⁵ indicate that about one-half of the people who receive questionnaires answer them, it was decided to send questionnaires to 23 college experts. Twenty-one of the 82 English composition and literature readers for the Advanced Placement Examination of the Educational Testing Service and two college English experts in Montana⁶ were contacted. This selection was made from the May, 1961, list of readers for the Educational Testing Service. In order to make the sample as nearly representative as possible, these experts were selected as representative of various areas of the United States and from institutions of various sizes. The writer personally knew two English experts who taught first year students in Montana colleges, and these were considered to be representative of college English experts of Montana.

A letter stating the purposes of this study and giving an explanation of what was to be done was sent to 23 college English experts, and a request for assistance⁷ was made. An itemization of the five criteria for the evaluation of English composition and a further sheet explaining

⁵Downie, N. M. and Heath, R. W., Basic Statistical Methods, pp. 103-104.

⁶Appendix B contains a complete list of the experts selected.

⁷See Appendix B for copies of letters sent.

each of the criteria⁸ were also sent. The college experts were asked to rank in order of importance the criteria listed.⁹

Nine of the 15 college English experts who answered were selected to rank the criteria. Six others withdrew as rankers because they did not feel that criteria for the evaluation of English composition should be ranked.

Pooley, although not requested to rank the five criteria, felt that criteria for the evaluation of English composition had value but that there was difficulty in getting teachers to use them, as shown in the following statement:

I would be pleased to help you with your study of composition but I regret to say that I have no first-hand information of any other analysis of the type you have in mind. Most of the composition scale studies were made in the 1920's and 1930's. There seems to have been very little activity in this connection in more recent years. I suppose the reason is that while composition scales are valuable in themselves, it is exceedingly difficult to get teachers to make use of them.¹¹

⁸Appendix B contains a copy of the itemization and definition of the criteria sent to the college experts for ranking.

⁹In the items listed, number one was to be the most important and number five was the least important. The experts were requested not to mark any items that they thought should not have been included.

¹⁰Appendix B gives the list of college experts, indicating those who withdrew as rankers.

¹¹Personal letter written by Dr. Robert C. Pooley, Head of the Department of Integrated Liberal Studies, University of Wisconsin, Madison, Wisconsin, and author of Teaching English Usage, 1946, and Teaching English Grammar, 1957. Letter written to the writer on May 14, 1962.

The statistical treatment of the numerical rankings of the criteria by nine college English experts with Kendall's coefficient of concordance, " W_s ," used to test the numerical rankings follows.

Statistical treatment of numerical rankings by college experts.

In order to reach an objective decision on the basis of the data collected in this study as to whether or not college English experts differ significantly in their ranking of the criteria, certain procedures were necessary prior to actual computation of the values in the statistical tests. It was necessary to state the null hypothesis (H_0) and the alternative hypothesis (H_1) for the ranking of the criteria by the college English experts. The null hypothesis was that there was no difference in the rankings of the criteria by the college English experts at the five per cent level. The alternative hypothesis was that the rankings of the criteria by college English experts were not the same. The rankings of the criteria by the nine college English experts is shown in Table 2.

Kendall's¹² coefficient of concordance, " W_s ," was used to test the null hypothesis because it is well adapted to ranked data. W_s ¹³ was found to equal .75 which is significant at the one per cent level, and the null hypothesis may be accepted; therefore, there is no significant difference among college experts throughout the United States in their ranking of

¹²Downie and Heath, op. cit., pp. 147-56.

¹³In Kendall's coefficient of concordance, " W_s ," perfect agreement is indicated by $W = 1$ and lack of agreement by $W = 0$.

TABLE 2. RANK GIVEN BY EACH OF THE NINE COLLEGE ENGLISH EXPERTS FOR IMPORTANCE OF EACH CRITERION

Criteria	Number of experts in each rank					Rank Order
	1	2	3	4	5	
Coherence and Logic	4	4	1			2
Development of Ideas	5	4				1
Diction			3	5	1	4
Emphasis				4	5	5
Organization Through Sentence Structure and Paragraphing			5	1	3	3

the criteria.¹⁴

Since the literature of leading authorities in the United States was used as the basis for the selection of criteria for the evaluation of English composition in this study, and since the college experts did have high composite agreement on the ranking of the criteria, the criteria used in this study were considered by the writer to have face or construct validity.

¹⁴In Appendix A, Table 20, the results of the statistical computations relative to agreement of college experts are presented.

Validation of Criteria by High School English Experts

The criteria were also validated by having high school English teachers rank the same points. The same method used to check the rankings of the criteria given by the college English experts was also used to check the rankings of the criteria given by the high school English teachers.

Selection of high school experts. Letters were sent to 187 high schools in Montana requesting the assistance of English teachers. These teachers were asked to rank the criteria and to grade 20 themes according to the criteria. Forty teachers¹⁵ from 30 different schools agreed to take part in the study. Of this number, 30 teachers from 26 different schools were selected. Of these 30 teachers, 22 ranked the criteria.¹⁶ The high school English teachers did not suggest using any other criteria for judging English composition. Their comments constituted constructive criticism of the scale sent to them for evaluation. An example of this type of criticism was that of MacDonald. She didn't believe that emphasis should be listed as a division and stated her belief as follows:

¹⁵Appendix C contains a complete list of high school teachers who agreed to participate in the study.

¹⁶For additional qualifications of high school English teachers who took part in the study, see Chapter 4, p. 57.

Even though I have given lower ratings on the emphasis than on any other division of the criteria, I do not think that it should be listed as a division. I believe that it should be included in the definition of Development of Ideas. It is my opinion that introduction and conclusion should be an addition to Organization, which should be an overall picture of the theme.¹⁷

Many of these high school English teachers felt that high schools should emphasize correct grammar, correct spelling, and proper word meaning.

Most of the high school English teachers felt that all the criteria were important or that some of the criteria should be combined. Some of them expressed the feeling that the criteria should have included mechanics as a specific item.

Statistical treatment of the numerical rankings by high school experts. Kendall's coefficient of concordance, "W," was used to determine the relationship among the rankings given the criteria by the high school experts, and this is shown in Table 3.

The null hypothesis of no difference in the ranking was tested in the same manner as with the rankings of the college experts. W was found to equal .65 which is significant at the one per cent level, and the null hypothesis may again be accepted.¹⁸

In the ranking of the criteria, the college English experts had agreement of $W = .75$, and the high school English experts had agreement

¹⁷These were comments written by Dorothy K. MacDonald, Whitefish High School, Whitefish, Montana, April, 1962, to the writer.

¹⁸In Appendix A, Table 19, the results of the statistical computations relative to agreement of high school English teachers are presented.

TABLE 3. RANK GIVEN BY EACH OF THE 22 HIGH SCHOOL ENGLISH EXPERTS FOR IMPORTANCE OF EACH CRITERION.

Criteria	Number of high school English experts in each rank					Rank Order
	1	2	3	4	5	
Coherence and Logic	6	13	3			2
Development of Ideas	14	6		2		1
Diction		2	7	10	3	4
Emphasis			3	2	17	5
Organization Through Sentence Structure and Paragraphing	2	2	8	8	2	3

of $W = .65$.¹⁹ With these significant figures, the criteria were found to be valid at the one per cent level. The rank given to each of the criteria by the college experts is presented in Table 2, and the rank given to each of the criteria by the high school experts is presented in Table 3.

Summary

Nine sources were selected as the basis from which criteria were determined for the evaluation of English composition. Examples of the

¹⁹See footnote 13, page 47, for interpretation of the significance of W .

sources used are the California Association of Teachers of English, the National Council of Teachers of English, and the College Entrance Examination Board. Important criteria for the evaluation of English composition were selected through an extensive examination of literature and were validated by college and high school English experts. These criteria were (1) Coherence and Logic, (2) Development of Ideas, (3) Diction, (4) Emphasis, and (5) Organization Through Sentence Structure and Paragraphing.

The criteria were ranked by nine college English experts and 22 high school English experts. The rankings of the experts were tested by Kendall's coefficient of concordance, " W_s ." W was found to equal .75 for the college experts, and W equaled .65 for the high school experts. In both cases, there was significant agreement at the one per cent level. Since the experts were in significant agreement in their rankings, face or construct validity was established. The composite agreement of the rankings in terms of importance by both the college and high school experts was (1) Development of Ideas, (2) Coherence and Logic, (3) Organization Through Sentence Structure and Paragraphing, (4) Diction, and (5) Emphasis.

The next step in the investigation was to establish reliability of the criteria. The rating of the themes to establish reliability for the criteria for the evaluation of English composition follows in Chapter 4.

CHAPTER IV

ESTABLISHMENT OF RELIABILITY OF THE CRITERIA FOR THE EVALUATION OF ENGLISH COMPOSITION

Reliability of the criteria was tested by having 30 high school English teacher experts correct the same themes according to the validated criteria.¹ In testing for reliability, an analysis of variance test was used to test the differences in the high school teacher experts' ratings of themes according to the criteria, and chi square tests were used to determine the significance of the ratings of the themes according to the criteria.

Four high schools in Montana were selected where twelfth grade students wrote themes² for this study. "Neighbors, Good and Bad--People" was chosen as the topic for the themes with the length of 200-300 words. Instructions³ were sent to the teachers to follow when administering the writing.

Every thirteenth theme was selected to be used from 256 received from four high schools. These themes were then graded by 30 teacher experts according to the validated criteria, and statistical analyses were

¹The validated criteria were (1) Coherence and Logic, (2) Development of Ideas, (3) Diction, (4) Emphasis, and (5) Organization Through Sentence Structure and Paragraphing.

²See Appendix C for copies of student themes.

³See Appendix C for a copy of the instructions which the teachers were to follow when administering the writing of the themes.

made of their numerical ratings. In the rating of each theme according to each criterion, teachers used a mark of "one" to denote poor, "two" for average, and "three" for better than average. As a check upon the ratings given the themes according to each criterion, over-all ratings were given. The numerical ratings of the themes according to each criterion and according to the over-all ratings were checked by statistical analyses. The selection of themes follows.

Selection of Themes Written by High School Students

In the selection of themes, certain procedures were followed. Instructions on how the themes were to be written were sent to the teachers who supervised the writing of the themes in four Montana high schools. A random selection of themes to be used in this study was made from those which were submitted.

Instructions for writing themes. "Neighbors, Good and Bad--People" was chosen as the topic, and the incident which was to be written about was to have taken place in one day or less. It was felt that the students' interest level was such that they would be motivated sufficiently to put forth their best efforts. This feeling that these students have an immense interest in the people about them developed from the writer's teaching experience at the twelfth grade level.

When deciding upon the length of each theme, certain questions arose: How long should it be in order that the theme contained enough

information, and how short should it be so that readers would not spend an excessive amount of time in evaluation? A good length had to take these things into consideration. The writer considered 200-300 words, chosen arbitrarily, as the length of themes.

A time limit was also set. Not more than one class period was to be utilized in which students were to write themes in order to give each student an equal chance and to avoid making the task burdensome for the teachers who supervised the writing.

How themes were selected. The methods used to gain the cooperation of the schools where the themes would be written and to actually obtain themes were determined. Four schools in which the writer personally knew the English teachers were selected, and contact was made with the individual teachers. These teachers⁴ agreed to supervise the writing of themes and working agreements with the writer were established. Agreement was reached that all the compositions were to be written during the same week in an effort to minimize differences due to regular instruction in the students' scheduled classes. Directions⁵ were sent to the four teachers who were selected to supervise the writing of themes. Two

⁴The four teachers who supervised the writing of themes were Mr. Neil McFadgen, Havre High School, Havre, Montana; Mr. Morgan Sherlock, Conrad High School, Conrad, Montana; Mr. Duane Hoynes, Fort Benton High School, Fort Benton, Montana; and Miss Gertrude Conwell, Sweet Grass County High School, Big Timber, Montana.

⁵Appendix C contains the directions sent to high school teachers who supervised the writing of the themes.

hundred and fifty six themes were completed and received by the writer.

Twenty themes were arbitrarily chosen as the number to be used in this study in order to simplify the judging process. The random sample of 20 themes was made by choosing every thirteenth theme from the 256 completed themes when they were not sorted in any particular way. The writer decided that 10 themes written by boys and 10 themes written by girls would be desirable to eliminate any sex variable in the statistical analyses. The selection was a completely random one, and it was found that 10 themes were written by boys and 10 themes were written by girls. The distribution of boys and girls by four schools was as follows:

Hayre High School--five by boys and six by girls

Conrad High School--two by boys and two by girls

Fort Benton High School--two by boys and none by girls

Sweet Grass County High School--one by a boy and two by girls

The next step in this study was the preparation of these compositions for mailing to 40 high school English teachers who were to rate the 20 themes according to the validated criteria and give over-all grades to the themes. Each of the themes was typed to eliminate prejudices that often occur due to paper size, handwriting style, and similar variables, and in order to simplify and minimize time in the evaluation. The following materials were mailed to the 40 high school English teachers who agreed to participate in the study: the typed themes, a copy of the rating sheets on which evaluation would be made, and a copy of directions

for judging. The establishment of reliability of the criteria follows.

Establishment of Reliability of the Criteria
by High School English Experts

Reliability of the five selected criteria for the evaluation of English composition would be established if the high school experts agreed in the ratings given each theme according to each criterion. The methods used in the selection of the experts, and the statistical treatment of their numerical ratings follow.

Selection of high school English experts. A form letter⁶ was sent to 187 high schools in Montana asking for teachers to judge themes according to specified criteria. These teachers were to have a Bachelor of Arts, Bachelor of Science, or higher degree, have a teaching major in English, and have at least two years experience in judging themes. It was also decided not to use more than three judges from any school since an over-balance in the judging could result if teachers who had worked closely together for a number of years graded the themes. A total of 40 instructors from 30 different schools agreed to participate. A second form letter⁷ accompanied by the criteria for evaluation of English composition, themes, directions, and grading sheets was mailed to the 40

⁶See Appendix C for a copy of the form letter sent to the 187 high schools in Montana.

⁷See Appendix C for materials sent to high school English experts.

teachers, and 30 completed ratings from 26 different schools were returned. Ten of these completed ratings were received through a follow-up letter sent to 20 teacher experts who had not returned the rating sheets. The ten ratings received after sending the follow-up letter were used as a check upon the first 20 ratings since the last group of 10 teachers had more time to consider the grade given each theme. Due to the "end of the year" period in which completion was to be made, 10 teacher experts were unable to finish the evaluation. The statistical treatment of the numerical ratings given each theme by each of 30 high school teacher experts with an analysis of variance test and chi square tests used to test the numerical ratings follows.

Statistical treatment of the numerical ratings of the high school English experts. In order to reach an objective decision on the basis of the data collected as to whether or not Montana high school English teacher experts differ significantly in evaluating themes according to validated criteria, certain procedures were necessary prior to actual computation of the values in the statistical tests. The preliminary procedures involved (1) the statement of the hypothesis and (2) the choice of the statistical techniques.

The first step in the statistical procedure was the statement of the null hypothesis (H_0) and the alternative hypothesis (H_1) for the evaluation of themes according to validated criteria. The null hypothesis was that there were no differences among high school English experts of Montana

in the evaluation of student themes for quality in the content at the five per cent level when using the validated criteria. The alternative hypothesis was that Montana high school English experts would differ in the evaluation of student themes for quality of content when using the validated criteria.

To reach an objective decision on the reliability of the scale used by the high school English experts in marking of the themes, it was necessary to test the following:

1. The differences among the experts in rating the themes when using the rating scale
2. The differences among the students in writing the themes
3. The homogeneity of the overall grades given by 20 high school English experts who corrected the themes
4. The homogeneity of the overall grades given by the other 10 high school English experts
5. The homogeneity of the grades for each of the criteria--the grades given by 20 high school English experts who corrected the themes
6. The homogeneity of the grades for each of the criteria--the grades given by the 10 other high school English experts
7. The homogeneity of the total markings of the five criteria by the first 20 high school English experts

An analysis of variance test⁸ was chosen to test for differences among the experts in rating the themes by the rating scale, and to test for differences in the writing of the themes among the students. To state

⁸Appendix A contains the analysis of variance calculations.

this in another way, the following questions can be asked: (1) Was there any difference in the rating of the themes by the individual experts?

(2) Was there a variation in the quality of the themes according to the way the experts rated them?

The chi square test was used to test the homogeneity of the overall grades given by the first 20 of the high school English experts and to test the homogeneity of the overall grades given by the last 10 high school English experts. After testing for homogeneity of the overall grades given by the first 20 and the last 10 high school English experts, the chi square test was used to test each criterion for homogeneity for the first 20 experts and then for the last 10 high school English experts. The chi square test was also used to test to see if there was homogeneity in the total markings of the five criteria by the first 20 high school English experts. The chi square method was used because the calculation procedures for homogeneity and chi square are exactly the same according to Cramer.⁹

When the 40 high school English experts agreed to participate in this study, they were sent six grading sheets together with 20 themes which were to be graded. One of the grading sheets was provided for each criterion and one grading sheet was provided for an overall grade.¹⁰

The scheme for grading the quality of themes when using the five criteria was that the mark of "one" was for poor, "two" was for average,

⁹Cramer, Harold, Mathematical Methods of Statistics, pp. 445-9.

¹⁰See Appendix C for copies of grading sheets.

and "three" was for better than average. The ratings for all criteria were combined for each theme for each expert. The lowest possible score on a composition by any expert was five since there were five criteria; the highest possible score was fifteen since three was the highest grade and there were five criteria.¹¹ From this information an analysis of variance table¹² was constructed.¹³

In the statistical analyses, an analysis of variance was used to test the significance of the differences among the teachers in rating the themes when using the criteria of the rating scale or stated in another way, the mean grades of teachers did not differ except by chance. It was found that the F ratio was 4.314 with 19 and 361 degrees of freedom which was a significant figure at the five per cent level. The null hypothesis was rejected and the alternative hypothesis that Montana high school English teachers would differ in the evaluation of student themes for quality of content when using the specified criteria was accepted.

An analysis of variance test was also used to test the differences among the students in writing the themes or stated in another way, the mean grades of the students did not differ except by chance. It was found that the F ratio of 15.375 with 19 and 361 degrees of freedom was a

¹¹Appendix A, Table 6, contains cumulative scores of teachers' grades.

¹²See Appendix A, Table 7, for analysis of variance table.

¹³Downie, N. M. and Hearsh, R. W., Basic Statistical Methods, pp. 445-9.

significant figure at five per cent level,¹⁴ and the null hypothesis was rejected. The alternative hypothesis that Montana high school English experts would differ in the evaluated criteria was accepted.

On the overall grading sheets¹⁵ received from 20 high school English experts, a grade of "one" was given as the lowest possible and "five" as the highest possible grade. The grades given for the ratings were added and compiled in a contingency table for the observed and expected frequencies from which the chi square value was obtained. For the 30 teacher experts, the value of chi square was obtained by the same process for the 20 experts who served as the original group as for the 10 remaining experts.

To test the null hypothesis that there was homogeneity of an overall grades given by 20 high school English experts marking 20 themes the chi square test was used. A contingency table¹⁶ was set up wherein the value of chi square was found to equal 114.57 with 76 degrees of freedom. This figure was significant at the five per cent level. The null hypothesis was rejected and the alternative hypothesis that Montana high school English experts would differ in the evaluation of student themes for quality in content when using the specified rating scale was accepted.

¹⁴See Appendix A, Table 7, for analysis of variance table.

¹⁵See Appendix C for a copy of the overall grading sheet.

¹⁶Appendix A contains Table 8 which shows the frequency of an overall grade given by 20 high school English teachers marking 20 themes.

The null hypothesis that there was homogeneity of an over-all grade given by the last 10 high school English experts marking 20 themes was tested by the chi square test. A contingency table¹⁷ was set up and the value of chi square was found to be 99.32 with 36 degrees of freedom. This was a significant figure at the five per cent level, and according to the "Distribution of Chi² Table,"¹⁸ the null hypothesis was rejected, and the alternative hypothesis that Montana high school English experts would differ in the evaluation of student themes for quality in content when using the validated criteria was accepted.

It was found that the experts did not agree in their grading of the 20 themes presented, according to the overall grades given. Since statistically significant figures did not show reliability of the criteria at the five per cent level, reliability when using the grading scale as a basis for grading could not be established.

To test the null hypothesis that there were no differences among high school English experts of Montana in evaluating student themes for quality in content when using the validated criteria each criterion was tested separately.

In testing each criterion separately, the chi square test was used.

¹⁷See Appendix A, Table 9, for the frequency of an overall grade given by the last 10 high school English teachers marking 20 themes.

¹⁸Lewis, Edward E., Methods of Statistical Analysis in Economics and Business, pp. 283 and 160-161.

A contingency table¹⁹ was set up for each criterion, and Table 4 shows the values of chi square with 38 degrees of freedom for each of the criteria for the first 20 experts. The null hypothesis that there were no differences among high school English experts of Montana in the evaluation of student themes for quality in the content when using the validated criteria was accepted at the one per cent level for "Organization Through Sentence Structure and Paragraphing." For the remaining criteria the chi square values were large, and the null hypotheses of no differences were rejected. The alternative hypothesis that Montana high school English experts would differ in the evaluation of student themes for quality in content when using the validated criteria was accepted.

TABLE 4. VALUES OF CHI SQUARE WITH 38 DEGREES OF FREEDOM FOR EACH OF THE CRITERIA FOR THE FIRST 20 EXPERTS

Criteria	Values of Chi Square
Coherence and Logic	80.698
Development of Ideas	71.53
Diction	72.386
Emphasis	60.301
Organization Through Sentence Structure and Paragraphing	54.076

The ratings of the 10 additional teachers were tested separately by

¹⁹Appendix A. Tables 10 through 14, inclusive, contain the frequency marking of each individual criterion for the first 20 teachers.

using the chi square method. A contingency table²⁰ was set up for the criteria and Table 5 shows the values of chi square with 18 degrees of

TABLE 5. VALUES OF CHI SQUARE WITH 18 DEGREES OF FREEDOM FOR EACH OF THE CRITERIA FOR THE 10 ADDITIONAL EXPERTS

Criteria	Values of Chi Square
Coherence and Logic	40.973
Development of Ideas	57.116
Diction	64.428
Emphasis	24.516
Organization Through Sentence Structure and Paragraphing	56.545

freedom for each of the criteria for the 10 additional experts. The null hypothesis that there were no differences among high school English experts of Montana in the evaluation of student themes for quality in the content when using the validated criteria was accepted at the five per cent level for "Emphasis." The remaining criteria had values of chi square that were large, and the null hypothesis was rejected. The alternative hypothesis that Montana high school English experts would differ in the evaluation of student themes for quality of content when using the validated criteria was accepted.

In testing the criteria individually, it was found that for the

²⁰Appendix A, Tables 15 through 19, inclusive, contain the frequency markings of each individual criterion for the last 10 experts.

first 20 experts the only significant agreement was for the criterion "Organization Through Sentence Structure and Paragraphing." For the next 10 experts the only significant agreement was for the criterion "Emphasis." In all other cases the alternative hypothesis that Montana high school English experts would differ in the evaluation of student themes for quality of content when using the validated criteria was accepted.

By combining the markings of the first 20 experts for all themes for all five criteria, the null hypothesis that there were no differences among high school English experts of Montana in the evaluation of student themes for quality in the content when using the validated criteria was checked. The scores each expert gave to all 20 themes for all the criteria were added to obtain a total marking for the combined criteria. In this marking "one" equaled poor, "two" equaled average, and "three" equaled better than average. There were five criteria for 20 themes marked by each expert for a total of 100 marks for each of the 20 experts. The null hypothesis that there was homogeneity of the total markings of the five criteria by the 20 high school English experts was tested by the chi square test. A contingency table²¹ was set up, and the value of chi square was found to be 256.997 with 38 degrees of freedom. This was a significant figure, and the null hypothesis was rejected. The alternative hypothesis that Montana high school English experts

²¹See Appendix A, Table 20, for contingency table.

differed in the evaluation of student themes for quality in content when using validated criteria was accepted.

When using the validated criteria, in all but two of the 15 tests made in checking the reliability of the grading scale for the evaluation of English composition the null hypothesis was rejected, and the alternative hypothesis that Montana high school English experts would differ in the evaluation of student themes was accepted. The two tests where the null hypothesis was accepted were for individual criteria; however, one criterion, "Organization Through Sentence Structure and Paragraphing," was in the sample of 20 experts, and the other criterion, "Emphasis," was in the sample of 10 experts. Because of the acceptance of the alternative hypothesis, that Montana high school English teacher experts would differ in the evaluation of student themes for quality in content when using the validated criteria, reliability for the criteria was not established.

The writer felt that since the criteria presented in this study were used more often than other criteria by The National Council of Teachers of English, by state English associations, by national testing services, and by nationally known experts in the field of English that these criteria should have been valid and reliable. The criteria did prove to be valid but not reliable. In addition, the teacher experts did not agree on the over-all grades given to the themes.

Since reliability was not found at the five per cent level when

testing the ratings given the 20 themes by 30 teacher experts according to the validated criteria or according to over-all grades, it is probable that individual differences among the teacher experts in their opinions of quality caused lack of reliability for the criteria even though they did agree in the ranking in terms of importance.

Summary

Reliability of the criteria for the evaluation of English composition was determined by the use of an analysis of variance test and chi square tests. The criteria were selected from nationally known sources in literature, and each of the 20 themes was corrected according to each criterion by each of 30 high school English experts.

In one check for reliability, the mean grades that teachers gave each theme according to each criterion were tested to determine if the grades differed significantly at the five per cent level. An analysis of variance test was used, and it was found that the F ratio was 4.314 with 19 and 361 degrees of freedom. The mean grades of teachers did differ at the five per cent level, and reliability of the criteria was not found.

In the second test for reliability of the criteria, the mean grades that students received on each theme according to each criterion were tested to determine if the grades differed significantly at the five per cent level. The same analysis of variance test as that used to test the mean grades of teachers was used to test the mean grades of students,

and it was found that the F ratio equaled 15.375 with 19 and 361 degrees of freedom. The mean grades that students received did differ at the five per cent level, and reliability of the criteria was not found.

In tests for reliability of the criteria for the evaluation of English composition that followed the use of the analysis of variance test, chi square tests were used. Each criterion was tested separately for reliability of the group of 20 teachers rating each theme according to each criterion and for the group of 10 teachers rating each theme according to each criterion. For "Organization Through Sentence Structure and Paragraphing" for the first group of 20 teachers, chi square was found to equal 54.076 with 38 degrees of freedom, and there was agreement at the five per cent level. For the first group of 20 teachers, agreement was not found at the five per cent level for any of the other criteria. For the criterion, "Emphasis," agreement was found at the five per cent level for the group of 10 teachers. Chi square was found to equal 24.516 with 18 degrees of freedom. For the second group of 10 teachers, agreement was not found at the five per cent level for any of the other criteria. Because agreement was for different criteria for each group of teachers, reliability of the criteria was not established.

In another test for reliability of the criteria, the chi square test was used to test for homogeneity of the total markings of the five criteria by the first group of 20 experts. Chi square was found to equal 256.007 with 38 degrees of freedom. Homogeneity did not exist at

the five per cent level, and reliability of the criteria was not found.

Reliability of the criteria for the evaluation of English composition was not found since there was not significant agreement at the five per cent level among Montana high school English experts in the evaluation of student themes according to the validated criteria. The criteria which were tested for reliability were (1) Coherence and Logic, (2) Development of Ideas, (3) Diction, (4) Emphasis, and (5) Organization Through Sentence Structure and Paragraphing.

The homogeneity of overall ratings given each theme by each teacher expert was tested by the chi square method. For the first 20 teachers chi square was found to equal 114.57 with 76 degrees of freedom, and homogeneity did not exist at the five per cent level. For the second group of 10 teachers, chi square was found to equal 99.32 with 36 degrees of freedom, and homogeneity did not exist at the five per cent level. The summary, conclusion, and recommendations appear in Chapter 5.

CHAPTER V

SUMMARY, CONCLUSION, AND RECOMMENDATIONS

The major purpose of this study was to develop a set of criteria for the evaluation of English composition. Literature was examined, and criteria¹ were carefully selected from nine sources² which were considered important because of their repeated occurrence and because of their presentation by state or nationwide authorities. Validity of the criteria was tested by using Kendall's coefficient of concordance, "W," and an analysis of variance test together with chi square tests were used to test for reliability of the criteria. Validity of the criteria was established by agreement at the one per cent level in the ranking of the criteria by college³ and high school⁴ experts. Twenty themes were randomly selected from 256 which were written by the seniors of four Montana high schools. In order to establish reliability of the validated criteria, statistical analyses were made of the numerical ratings of 20 themes by 30 teacher experts from 26 Montana high school English staffs. These 30 experts also gave overall ratings to the themes, and these ratings were also tested to

¹The criteria selected were (1) Coherence and Logic, (2) Development of Ideas, (3) Diction, (4) Emphasis, and (5) Organization Through Sentence Structure and Paragraphing.

²See Chapter 2, page 39, footnote #1, for a list of the nine sources.

³See Appendix B for a list of the college experts.

⁴See Appendix C for a list of the high school experts.

determine reliability.

Summary

During the first two decades of the twentieth century, comparison scales were devised by selecting representative themes for each letter grade and using these themes as a basis for other grading. With the aid of the comparison scales, some consistency in grading English composition seemed evident, but when the early scale developers checked their scales for validity and reliability, they found that correlation coefficients were low.

The first reported attempts to establish criteria for the evaluation of English compositions took place during the period following World War I. From that time until 1962, numerous scales were developed which used criteria in the evaluation process with spelling, punctuation, and organization emphasized as the most important criteria, but validity and reliability were not established.

In this study, criteria were selected from nine pamphlets or articles of national importance which were published from 1951 to 1962 by The National Council of Teachers of English, state associations of teachers of English, national testing services, and individuals of national importance in the field of English, and these criteria were not tested for reliability of validity. It was believed by the writer that criteria which were used most often by these authorities in literature should prove to be

valid and reliable.

The validated criteria used in this study were ranked in terms of importance by nine college English experts and 22 high school English experts. The significance of the numerical ranking was determined by Kendall's coefficient of concordance, "W". W equaled .75 for the college experts, and W equaled .65 for the high school experts. Agreement was found at the one per cent level for both groups, and validity was established. The agreement in the ranking of the criteria was as follows: (1) Development of Ideas, (2) Coherence and Logic, (3) Organization Through Sentence Structure and Paragraphing, (4) Diction, and (5) Emphasis.

Twenty themes were selected from 256 which were submitted from four Montana high schools. The topic of the themes was "Neighbors, Good and Bad--People," and each student wrote 200-300 words. Reliability of the validated criteria was measured by testing the ratings given the 20 themes according to each criterion by 30 high school English experts. Reliability of the overall grades given each theme by each high school teacher expert was also determined.

Reliability of the criteria was not found at the five per cent level when an analysis of variance test was made to test the mean grades that teachers gave each theme according to each criterion and to test the mean grades that students received on each theme according to each criterion. Reliability was found in two of the chi square tests which were

made. Chi square was found to equal 54.076 with 38 degrees of freedom, and reliability was found at the one per cent level for the first 20 teachers for "Organization Through Sentence Structure and Paragraphing." Chi square was found to equal 24.516 with 18 degrees of freedom, and reliability was found at the one per cent level for the last 10 teachers for "Emphasis." Since the agreement was for a different criterion in each group, this was not conclusive evidence of reliability for the two criteria. Chi square was found to equal 256.997 with 38 degrees of freedom when combining the five criteria and testing the numerical rankings. There was not agreement at the five per cent level, and reliability was not established. An overall rating was given each theme when not using the criteria. Reliability was not established at the five per cent level for the overall ratings since chi square was found to equal 114.57 with 76 degrees of freedom.

Literature revealed that comparison scales and criteria for the evaluation of English composition did not prove to be valid and reliable when tested. In this study, the criteria of (1) Coherence and Logic, (2) Development of Ideas, (3) Diction, (4) Emphasis, and (5) Organization Through Sentence Structure and Paragraphing were found to be valid, but reliability was not found.

Conclusion

Perhaps the only conclusion justified by this study is that,

although teachers of English composition may feel that criteria are important in evaluating themes, there is no evidence of consistency in the employment of such criteria.

Recommendations

Since it is probable that reliability of the criteria would be found if themes were rated by a group of English theme raters who worked closely together, it is recommended that in a future study use be made of the criteria which were validated in this study. Although the writer defined each criterion for the use of the teacher experts, he believes that reliability would be established for the criteria if high school English experts from the same high school were oriented for theme grading according to the criteria before rating themes. It is the writer's belief that if the validated criteria were treated individually and defined with specific points agreed upon by the raters under each criterion, then reliability of the criteria would be established. If English teachers would meet on a regional basis and agree on the usage of each of the valid criteria found in this study, then application of this instrument would most likely show reliability.

LITERATURE CONSULTED

LITERATURE CONSULTED

- American Studies Association, and others, The Basic Issues in the Teaching of English, 15 pp., (Published as a supplement to College English 21:1-15, October, 1959.).
- Anderson, Harold A., and Traxler, Arthur B., "Reliability of an Essay Test in English," 1935 (As related in Anderson, H. A., and Traxler, A. B., "The Reliability of the Reading of an English Essay Test: A Second Study," School Review, 48:521-530, September, 1940.).
- California Association of Teachers of English, A Scale for Evaluation of High School Student Essays, 1960, 32 pp.
- "College Bowl," "Why I Want to Go to College," (General Electric Series), produced by ABC Television Network, broadcast by KXLF T.V., Butte, Montana, 3:30 p.m., February 25, 1962, Allen Ludden, M. C.
- College Entrance Examination Board, "Report of the 1961 Western Advanced Placement English Conference," Occidental College, Los Angeles, California, June 22-24, 1961, 16 pp.
- Coward, Ann F., "A Comparison of Two Methods of Grading English Composition," Journal of Educational Research, 46:81-93, October, 1952.
- Cramer, Harold, Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey, 1957, 575 pp.
- Dewey, John, The School and Society, University of Chicago Press, Chicago, Illinois, 1900, 164 pp.
- Diederich, Paul B., Notes on Grading Essays, Educational Testing Service, Princeton, New Jersey, November 22, 1955, 6 pp.
- Downie, N. M., and Heath, R. W., Basic Statistical Methods, Harper and Brothers Publishers, New York, 1959, 289 pp.
- Dusel, W. J., "How Should Student Writing Be Judged?" English Journal, 46:265, May, 1957.
- Educational Testing Service, A Brief--Cooperative Sequential Tests of Educational Progress, Princeton, New Jersey, 1957, 24 pp.
- Estrin, H. A., "How Do You Grade a Composition?" College Composition and Communication, 12:234-235, December, 1961.

Grose, L. M., Miller, Dorothy, and Steinberg, E. R., editors, Suggestions for Evaluating Junior High School Writing, National Council of Teachers of English, Champaign, Illinois, n.d., 49 pp.

Hillegas, Milo B., "A Scale for the Measurement of Quality in English Composition by Young People," College Teachers Record, 13:1-54, September, 1912.

Huddleston, Edith M., "Measurement of Writing Ability at the College Entrance Level: Objective vs. Subjective Testing Techniques," Journal of Experimental Education, 22:204, March, 1954.

Hudelson, Earl, English Composition, Its Aims, Methods, and Measurement, Public School Publishing Co., Bloomington, Ill., 1923, 173 pp.

Hutchins, Robert Maynard, No Friendly Voice, University of Chicago Press, Chicago Ill., 1936, 197 pp.

Huxtable, Zelma L., "Criteria for Judging Thought Content in Written Composition," Journal of Educational Research, 19:190, March, 1929.

Illinois Association of Teachers of English, Illinois English Bulletin, vol. 40, no. 6, "Evaluating Ninth-Grade Themes," Urbana, Ill., March, 1953, 35 pp.

Illinois Association of Teachers of English, Illinois English Bulletin, vol. 40, no. 7, "Evaluating Twelfth-Grade Themes," Urbana, Ill., April, 1953, 50 pp.

Indiana Council of Teachers of English, Indiana English Leaflet, vol. 3, no. 1, "Standards for Written English in Grade 12," October, 1960, 24 pp.

Koclanes, T. A., "Can We Evaluate Compositions?" English Journal, 50:250-257, April, 1961.

Lewis, Edward E., Methods of Statistical Analysis of Economics and Business, Houghton Mifflin Company, Boston, Mass., 1953, 686 pp.

Lynch, James J., "The Achievement of Excellence in Composition," The High School Journal, 45:14-20, October, 1961.

McKean, Keith, "If the Shoe Fits," College English, 8:255-261, February, 1947.

- Miller, Joseph Washburn, An Analysis of Freshman Writing at the Beginning and End of a Year's Work in Composition, Doctoral dissertation, University of Minnesota, Minneapolis, 1958, 208 pp., microfilm.
- National Council of Teachers of English, "Ideaform Paper," Champaign, Illinois, 1955, 2 pp.
- Noyes, E. S., "Recent Trends in the Comprehensive Examination in English," Educational Record (Supplement 13) 21:118, January, 1940.
- Rice, J. M., "The Results of a Test in Language," The Forum, 35:269-273, October, 1903.
- Rice, J. M., "The Need of a New Basis in Education," The Forum, 35:443, January, 1904.
- Ruoff, James, "Evaluating Student Essays in Literature: A Plea for Objective Criteria," College English, 22:35-37, October, 1960.
- Stewart, Marietta, "A Scale for Measuring the Quality of Conventional News Stories in High School Journalism," English Journal, 23: 209-215, March, 1934.
- Thomas, Ednah S., Evaluating Student Themes, University of Wisconsin Press, Madison, Wisconsin, 1955, 39 pp.
- Thorndike, Edward L., "A Scale for Merit in English Writing by Young People," Journal of Educational Psychology, 2:361-368, September, 1911.
- Trabue, M. R., "Supplementing the Hillegas Scale," Teachers College Record, 18:51-84, January, 1917.
- Travers, R. M. W., "A Review of Procedures for the Evaluation of the Outcomes of Teaching English," Journal of Experimental Education, 17: 325-333, December, 1948.
- Tressler, J. C., and Christ, Henry I., English in Action, Course Two, D. C. Heath, Boston, Mass., 1960, 500 pp.
- University of the State of New York, The State Education Department, Suggestions on the Rating of Regents Examination Papers in English, n.d., 17 pp.

- Vickery, K. F., The Use of Objective Measurement of Writing Techniques, State Teachers College, Dickinson, North Dakota, September, 1959, 9 pp.
- Ward, W. S., editor, "Principles and Standards in Composition for Kentucky High Schools and Colleges," Kentucky English Bulletin, vol. 6, no. 1, Fall, 1956, 72 pp.
- Wiseman, Stephan, "Grammar School Selection: Marking an Essay Reliably," Times Educational Supplement, 1771:231, April, 1949.
- Wolfe, Don M., Geyer, Ellen M., and others, Enjoying English, L. W. Singer; Syracuse, New York, 1955, 105 pp.

Appendix A
Statistical Computations

Analysis of Variance

TABLE 6. CUMULATIVE SCORES OF TEACHER'S GRADING ON EACH THEME FOR COMBINED CRITERIA*

Themes	Ratings of 20 Teachers																				Total ($\sum X$)	$\sum x^2$
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
1	8	7	6	9	13	15	8	9	11	7	6	5	7	9	10	6	5	10	6	6	163	1463
2	6	5	7	8	7	10	5	8	5	6	5	6	7	6	6	5	5	8	6	9	130	886
3	6	7	8	9	7	5	8	7	5	5	5	10	5	11	5	6	6	11	7	10	143	1105
4	10	12	10	13	15	15	14	15	9	13	7	12	9	14	5	12	10	11	13	15	234	2888
5	9	15	14	13	15	15	15	7	12	15	11	10	14	14	10	7	11	12	12	12	243	3079
6	8	10	12	8	11	15	10	12	11	13	7	14	10	7	10	13	8	9	6	13	207	2265
7	5	8	10	6	13	10	9	10	7	7	5	12	5	12	5	12	6	8	8	15	173	1669
8	6	7	8	7	7	5	6	6	6	7	5	11	7	11	5	6	6	6	5	13	140	1072
9	11	15	9	14	14	15	9	9	14	11	10	15	13	6	10	9	6	10	9	11	220	2572
10	9	10	9	13	11	5	13	9	6	10	10	6	7	12	9	9	5	11	11	13	188	1890
11	11	12	7	13	7	15	9	13	14	10	13	8	11	11	10	8	10	8	10	11	211	2327
12	10	14	12	13	13	15	14	13	15	12	14	7	12	6	15	7	14	8	7	11	232	2866
13	9	8	5	6	11	10	5	9	9	12	10	5	6	15	15	5	11	12	5	8	176	1748
14	12	14	15	13	15	15	14	13	11	13	14	15	15	12	15	6	13	13	12	15	265	3597
15	9	11	10	15	11	15	10	13	13	14	10	11	11	9	14	13	8	11	11	10	229	2701
16	6	5	9	8	7	10	5	6	5	7	5	5	5	7	10	5	5	8	5	10	133	953
17	8	10	13	9	7	5	8	14	6	7	5	8	6	6	5	10	8	7	5	14	161	1453
18	7	15	10	15	14	15	14	15	9	10	8	13	13	12	10	13	9	10	11	14	237	2935
19	8	11	8	9	13	10	7	15	8	7	5	9	7	7	10	5	7	9	9	12	176	1670
20	9	9	10	9	6	5	5	12	7	7	5	7	6	9	10	6	11	11	7	13	164	1458
$\sum X =$	167	205	192	210	217	225	188	215	183	193	160	189	176	196	189	163	164	193	165	235		
$\sum x^2 =$	1465	1972	2557	1998	1881	1480	1754	2013	1494	1501												
	2303	2378	2875	2493	2041	1999	2090	1499	1929	2875												
																					$G = 3825$	$\sum y^2 = 40597$

*Poor = 1, average = 2, and better than average = 3.

Analysis of Variance--Calculation Procedures

To test the hypothesis that all mean grades of teachers are equal (to test the difference among the teachers in rating the items of the rating scale) calculate:

$$F = \frac{\text{sum of mean square--teachers (sst)}}{\text{sum of mean square--error (sse)}}$$

To test the hypothesis that mean scores of students are equal (to test the difference among the students in writing the themes) calculate:

$$F = \frac{\text{sum of mean square--students (sssu)}}{\text{sum of mean square--error (sse)}}$$

TABLE 7. ANALYSIS OF VARIANCE--F RATIOS CALCULATED FROM SUM OF SQUARES, DEGREES OF FREEDOM, AND MEAN SQUARES

Source of Variation	Sum of Squares	df	Mean Square	F
Teachers	448.29	19	23.594	4.314
Students	1597.63	19	84.085	15.375
Error	1974.52	361	5.469	
Total	4020.44	399	10.076	

Analysis of Variance--Calculation Procedures (continued)

Code for Table 7 and Calculation for Analysis of Variance

S.S.T. = Sum of Squares--Teachers
 S.S.Su. = Sum of Squares--Students
 S.S.E. = Sum of Squares--Error
 T = Total Sum of Squares
 sst = sum of mean square--teachers
 sssu = sum of mean square--students
 t = total sum of mean square

$\sum Y^2$ = See Table 6
 G = See Table 6

$$T = \sum Y^2 - \frac{G^2}{400}$$

$$T = 40597 - \frac{(3825)^2}{400}$$

$$T = 4020.44$$

$$t = \frac{T}{df}$$

$$t = \frac{4020.44}{399}$$

$$t = 10.076$$

$$S.S.T. = \frac{\sum_{i=1}^{20} T_i^2}{20} - \frac{G^2}{400}$$

$$S.S.T. = \frac{740497}{20} - 36,576.56$$

$$S.S.T. = 448.29$$

$$sst = \frac{S.S.T.}{df}$$

$$sst = \frac{448.29}{19}$$

$$sst = 23.594$$

Analysis of Variance--Calculation Procedures (continued)

$$S.S.Su. = \frac{\sum_{i=1}^{20} S_i^2}{20} - \frac{G^2}{400}$$

$$S.S.Su. = \frac{763483}{20} - 36,576.56$$

$$S.S.Su. = 1597.63$$

$$ssu = \frac{S.S.Su.}{df} \quad ssu = \frac{1597.63}{19} \quad ssu = 84.085$$

$$S.S.E. = T - S.S.T. - S.S.Su.$$

$$S.S.E. = 4020.44 - 448.29 - 1597.63$$

$$S.S.E. = 1974.52$$

$$sse = \frac{S.S.E.}{df} \quad sse = \frac{1974.52}{361}$$

$$sse = 5.469$$

$$F = \frac{sst}{sse} \quad F = \frac{23.594}{5.469} \quad F = 4.314^*$$

$$F = \frac{ssu}{sse} \quad F = \frac{84.085}{5.469} \quad F = 15.375$$

*For interpretation of "F ration," see Downie, N. M. and Heath, R. W., Basic Statistical Methods, Table V, pp. 268-273.

χ^2

(Chi Square)

TABLE 8. FREQUENCY OF OVERALL GRADES* GIVEN BY 20 HIGH SCHOOL ENGLISH EXPERTS MARKING 20 THEMES

Rank	Observed Frequencies (O)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
(1)	5	4	2	4	3	1	6	3	6	0	9	4	2	4	6	4	4	0	5	1	73
(2)	5	3	5	5	4	3	7	2	3	7	3	4	9	5	4	7	9	7	7	1	100
(3)	7	7	8	4	2	6	4	8	7	3	4	6	4	5	6	5	6	12	6	5	115
(4)	3	5	4	5	4	7	3	4	3	5	2	4	3	4	2	3	1	1	2	7	72
(5)	0	1	1	2	7	3	0	3	1	5	2	2	2	2	2	1	0	0	0	6	40
Total																					400

Rank	Expected Frequencies** (E)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
(1)	3.65	(Expected frequencies are the same in each rank,																			73
(2)	5.00	(1) through (5), for teacher experts one through																			100
(3)	5.75	20.)																			115
(4)	3.60																				72
(5)	2.00																				40
Total																					400

Key: df = degrees of freedom
 r = the number of rows in the contingency table
 c = the number of columns in the contingency table

$$df = (r-1)(c-1)$$

$$df = (4)(19)$$

$$df = 76$$

$$\chi^2 = \frac{(O - E)^2}{E}$$

$$\chi^2 = 114.57$$

*In ranking, (1) equaled the lowest possible grade and (5) equaled the highest possible grade.

**Downie; N. M. and Heath, R. W., Basic Statistical Methods, p. 150, can be seen for explanation of method.

TABLE 9. FREQUENCY OF OVERALL GRADES* GIVEN BY THE LAST 10 HIGH SCHOOL ENGLISH EXPERTS MARKING 20 THEMES

Rank	Observed Frequencies (O)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
(1)	2	0	3	0	0	0	4	6	4	13	32
(2)	3	6	3	1	6	0	3	7	6	5	40
(3)	7	10	6	13	7	14	7	1	5	2	72
(4)	4	2	2	5	5	5	3	3	3	0	32
(5)	4	2	6	1	2	1	3	3	2	0	24
Total											200

Rank	Expected Frequencies (E)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
(1)	3.2										32
(2)	4.0	(Expected frequencies are the same in each rank, (1) through (5), for experts one through 10.)									40
(3)	7.2										72
(4)	3.2										32
(5)	2.4										24
Total											200

$$df = (r-1)(c-1)$$

$$df = (4)(9)$$

$$df = 36$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 99.32$$

*In ranking, (1) equaled the lowest possible grade and (5) equaled the highest possible grade.

TABLE 10. FREQUENCY OF GRADES FOR THE FIRST 20 EXPERTS WHEN THEY GRADED CRITERION NUMBER ONE--"COHERENCE AND LOGIC"

Rank	Observed Frequencies (O)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	7	4	3	7	8	5	6	4	11	7	11	9	7	5	6	11	9	2	11	3	136
2 Ave.	10	8	13	6	3	5	9	5	6	7	6	5	9	7	10	5	9	14	8	6	151
3 Good	3	8	4	7	9	10	5	11	3	6	3	6	4	8	4	4	2	4	1	11	113
	Total																				400

Rank	Expected Frequencies (E)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	6.8																				136
2 Ave.	7.55																				151
3 Good	5.65																				113
	Total																				400

$$df = (r-1)(c-1)$$

$$df = (2)(19)$$

$$df = 38$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 80.689$$

TABLE 11. FREQUENCY OF GRADES FOR THE FIRST 20 EXPERTS WHEN THEY GRADED CRITERION NUMBER TWO--"DEVELOPMENT OF IDEAS"

Rank	Observed Frequencies (O)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	10	5	4	6	8	5	5	4	9	5	12	6	7	5	6	9	10	2	10	1	129
2 Ave.	10	9	14	6	4	5	9	7	6	11	5	8	6	5	10	7	8	16	9	9	164
3 Good	0	6	2	8	8	10	6	9	5	4	3	6	7	10	4	4	2	2	1	10	107
	Total																				400

Rank	Expected Frequencies (E)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	6.45	(Expected Frequencies are the same in each rank, (1) through (20), for experts one through 20.)																			129
2 Ave.	8.20																				164
3 Good	5.35																				107
	Total																				400

$$df = (r-1)(c-1)$$

$$df = (2)(19)$$

$$df = 38$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 71.53$$

TABLE 12. FREQUENCY OF GRADES FOR THE FIRST 20 EXPERTS WHEN THEY GRADED CRITERION NUMBER THREE--"DICTION"

Rank	Observed Frequencies (O)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	4	6	5	1	2	5	11	1	9	7	8	7	12	9	6	11	8	8	8	3	131
2 Ave.	15	7	12	13	12	5	5	10	5	8	11	9	6	7	10	6	9	10	7	11	178
3 Good	1	7	3	6	6	10	4	9	6	5	1	4	2	4	4	3	3	2	5	6	91
	Total																				400

Rank	Expected Frequencies (E)																				Total																				
	Ranking of 20 Experts																																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20																					
1 Poor	(Expected frequencies are the same in each rank, (1) through (3), for experts one through 20.)																				131																				
2 Ave.																					6.55	8.90	4.55																		178
3 Good																																									91
	Total																				400																				

$$df = (r-1)(c-1)$$

$$df = (2)(19)$$

$$df = 38$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 72.386$$

TABLE 13. FREQUENCY OF GRADES FOR THE FIRST 20 EXPERTS WHEN THEY GRADED
CRITERION NUMBER FOUR--"EMPHASIS"

Rank	Observed Frequencies (0)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	6	7	6	2	4	5	7	7	7	7	12	7	8	6	6	10	9	3	9	2	130
2 Ave.	14	9	11	11	7	5	9	9	7	10	5	10	9	10	11	8	10	15	11	9	190
3 Good	0	4	3	7	9	10	4	4	6	3	3	3	3	4	3	2	1	2	0	9	80
	Total																				400

Rank	Expected Frequencies (E)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	6.5																				130
2 Ave.	9.5																				190
3 Good	4.0																				80
	Total																				400

$$df = (r-1)(c-1)$$

$$df = (2)(19)$$

$$df = 38$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 60.301$$

TABLE 14. FREQUENCY OF GRADES FOR THE FIRST 20 EXPERTS WHEN THEY GRADED CRITERION NUMBER FIVE--"ORGANIZATION THROUGH SENTENCE STRUCTURE AND PARAGRAPHING"

Rank	Observed Frequencies (O)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	10	6	6	7	3	5	7	8	6	8	9	6	7	9	6	13	10	4	8	2	140
2 Ave.	10	6	10	8	7	5	8	6	9	5	9	9	10	7	10	3	8	14	8	8	160
3 Good	0	8	4	5	10	10	5	6	5	7	2	5	3	4	4	4	2	2	4	10	100
	Total																				400

Rank	Expected Frequencies (E)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	7.0	(Expected frequencies are the same in each rank, (1) through (20), for experts one through 20.)																			140
2 Ave.	8.0																				160
3 Good	5.0																				100
Total																					400

$$df = (r-1)(c-1)$$

$$df = (2)(19)$$

$$df = 38$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 54.076$$

Distribution of χ^2 Beyond 30 Degrees of Freedom

It was necessary to calculate χ^2 for 38 degrees of freedom at the 1% and 5% levels since χ^2 tables do not ordinarily show distribution beyond 30 degrees of freedom. The formula for these procedures and the calculations are as follows:

$$\sqrt{2\chi^2} - \sqrt{2n-1} = \text{formula for space under normal curve for correct percent level as given in the normal table}^*$$

At the 1% level, χ^2 was calculated as follows for 38 degrees of freedom.

$$\sqrt{2\chi^2} - \sqrt{2n-1} = 2.326 \quad \sqrt{2\chi^2} = 10.92$$

$$\sqrt{2\chi^2} - \sqrt{75} = 2.326 \quad 2\chi^2 = 119.25$$

$$\sqrt{2\chi^2} - 8.66 + 2.326 \quad \chi^2 = 59.63$$

At the 5% level, χ^2 was calculated as follows for 38 degrees of freedom:

$$\sqrt{2\chi^2} - \sqrt{2n-1} = 1.645 \quad \sqrt{2\chi^2} = 10.305$$

$$\sqrt{2\chi^2} - \sqrt{75} = 1.645 \quad 2\chi^2 = 104.19$$

$$\sqrt{2\chi^2} - 8.66 + 1.645 \quad \chi^2 = 52.09$$

*Lewis, Edward E., Methods of Statistical Analysis in Economics and Business, pp. 283 and 160-161.

TABLE 15. FREQUENCY OF GRADES FOR THE LAST 10 EXPERTS WHEN THEY GRADED CRITERION NUMBER ONE--"COHERENCE AND LOGIC"

Rank	Observed Frequencies (O)										Total
	Rankings of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	8	8	7	5	11	1	7	4	3	8	62
2 Ave.	5	10	9	12	9	10	7	8	12	10	92
3 Good	7	2	4	3	0	9	6	8	5	2	46
	Total										200

Rank	Expected Frequencies (E)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	6.2										62
2 Ave.	9.2										92
3 Good	4.6										46
	Total										200

$$df = (r-1)(c-1)$$

$$df = (9)(2)$$

$$df = 18$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 40.073$$

TABLE 16: FREQUENCY OF GRADES FOR THE LAST 10 EXPERTS WHEN THEY GRADED
CRITERION NUMBER TWO--"DEVELOPMENT OF IDEAS"

Rank	Observed Frequencies (O)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	8	9	6	18	4	4	8	6	3	4	70
2 Ave.	6	9	10	2	12	5	5	6	15	9	79
3 Good	6	2	4	0	4	11	7	8	2	7	51
	Total										200

Rank	Expected Frequencies (E)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	7.0										70
2 Ave.	7.9										79
3 Good	5.1										51
	Total										200

$$df = (r-1)(c-1)$$

$$df = (9)(2)$$

$$df = 18$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 57.116$$

TABLE 17. FREQUENCY OF GRADES FOR THE LAST 10 EXPERTS WHEN THEY GRADED CRITERION NUMBER THREE--"DICTION"

Rank	Observed Frequencies (O)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	5	8	3	20	5	1	2	3	5	8	60
2 Ave.	11	9	12	0	11	13	9	13	13	9	100
3 Good	4	3	5	0	4	6	9	4	2	3	40
	Total										200

Rank	Expected Frequencies (E)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	6.0										60
2 Ave.	10.0										100
3 Good	4.0										40
	Total										200

$$df = (r-1)(c-1)$$

$$df = (9)(2)$$

$$df = 18$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 64.428$$

TABLE 18. FREQUENCY OF GRADES FOR THE LAST 10 EXPERTS WHEN THEY GRADED CRITERION NUMBER FOUR--"EMPHASIS"

Rank	Observed Frequencies (O)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	3	8	7	11	3	2	9	7	3	7	60
2 Ave.	12	8	11	9	14	14	8	9	10	11	106
3 Good	5	4	2	0	3	4	3	4	7	2	34
	Total										200

Rank	Expected Frequencies (E)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	6.0										60
2 Ave.	10.6										106
3 Good	3.4										34
	Total										200

$$df = (r-1)(c-1)$$

$$df = (9)(2)$$

$$df = 18$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 24.516$$

TABLE 19. FREQUENCY OF GRADES FOR THE LAST 10 EXPERTS WHEN THEY GRADED CRITERION NUMBER FIVE--"ORGANIZATION THROUGH SENTENCE STRUCTURE AND PARAGRAPHING"

Rank	Observed Frequencies (0)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	9	8	5	20	2	3	3	3	4	7	64
2 Ave.	5	10	13	0	14	10	11	14	12	7	96
3 Good	6	2	2	0	4	7	6	3	4	6	40
	Total										200

Rank	Expected Frequencies (E)										Total
	Ranking of 10 Experts										
	1	2	3	4	5	6	7	8	9	10	
1 Poor	6.4										64
2 Ave.	9.6										96
3 Good	4.0										40
	Total										200

$$df = (r-1)(c-1)$$

$$df = (9)(2)$$

$$df = 18$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 56.545$$

TABLE 20. FREQUENCY OF GRADES FOR THE TOTAL MARKINGS OF THE FIVE CRITERIA BY THE FIRST 20 HIGH SCHOOL EXPERTS

Rank	Observed Frequencies (O)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	37	28	24	23	25	25	38	24	42	34	52	35	41	34	30	54	46	19	46	11	666
2 Ave.	59	39	60	44	33	25	40	37	33	41	36	41	40	36	51	29	44	69	43	43	843
3 Good	4	33	16	33	42	50	24	39	25	25	12	24	19	30	19	17	10	12	11	46	491
	Total																				2000

Rank	Expected Frequencies (E)																				Total
	Ranking of 20 Experts																				
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1 Poor	33.30																				666
2 Ave.	42.15																				843
3 Good	24.55																				491
	Total																				2000

$$df = (r-1)(c-1)$$

$$df = (2)(9)$$

$$df = 18$$

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

$$\chi^2 = 256.977$$

Kendall's Coefficient of Concordance, W

TABLE 21. CALCULATION OF THE COEFFICIENT OF CONCORDANCE, THE DATA CONSISTING OF THE RANKING OF THE FIVE CRITERIA BY MONTANA HIGH SCHOOL ENGLISH EXPERTS

Criteria	Ranking by 22 Experts																						Sum of Ranks	Order	D	D ²
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22				
1	2	2	2	1	2	1	1	2	2	2	2	2	2	1	2	1	3	1	3	3	2	2	41	2	25	625
2	1	1	1	2	1	2	2	4	1	1	1	1	4	2	1	2	1	3	1	1	1	1	35	1	31	961
3	3	4	4	3	3	4	3	3	3	4	4	4	5	5	4	5	4	2	2	4	4	3	80	4	14	196
4	5	5	5	4	5	5	5	5	5	5	3	5	3	3	5	4	5	5	5	5	5	5	102	5	36	1296
5	4	3	3	5	4	3	4	1	4	3	5	3	1	4	3	3	2	4	4	2	3	4	72	3	6	36
																							330			3114

Key: D = Difference of the sum of the ranks
 N = Number of criteria
 m = Number of judges

$$D = \frac{\sum \text{Total}}{\text{No. of criteria}}$$

$$D = \frac{330}{5}$$

$$D = 66$$

$$N = 5$$

$$m = 22$$

$$\sum D = 330 \quad \sum D^2 = 3114$$

$$W = \frac{12 \sum D^2}{m^2(N)(N^2 - 1)}$$

$$W = \frac{12 (3114)}{480 (5) (25 - 1)}$$

$$W = .65$$

TABLE 22. CALCULATION OF THE COEFFICIENT OF CONCORDANCE, THE DATA CONSISTING OF THE RANKING OF THE FIVE CRITERIA BY NINE COLLEGE ENGLISH EXPERTS

Criteria	Ranking by Nine English Experts									Sum of Ranks	Order	D	D ²
	1	2	3	4	5	6	7	8	9				
1	2	1	2	2	2	1	3	1	1	15	2	12	144
2	1	2	1	1	1	2	1	2	2	13	1	14	196
3	4	5	4	4	3	4	4	3	3	34	4	7	49
4	5	4	5	5	4	5	5	4	4	41	5	14	196
5	3	3	3	3	5	3	2	5	5	32	3	5	25
	$\Sigma = 135$												$\Sigma = 610$

Key: D = Difference of the sum of the ranks
 N = Number of criteria
 m = Number of judges

$$D = \frac{\Sigma \text{ Total}}{\text{No. of criteria}}$$

$$D = \frac{135}{5}$$

$$D = 27$$

$$N = 5$$

$$m = 9$$

$$W = \frac{12 \Sigma D^2}{m^2 (N) (N^2 - 1)}$$

$$W = \frac{12 (610)}{81(5) (25 - 1)}$$

$$W = .75$$

Appendix B

Materials Pertinent to the College Expert Ratings of the
Scale for Grading English Composition

College English Experts Contacted

The following list names the readers of May 1961 for the Advanced Placement Examination of the Educational Testing Service who were contacted in this study.

1. Armstrong, Ray L. **
Lehigh University
Bethlehem, Pennsylvania
2. Arp, Thomas R. **
Bowdoin College
Brunswick, Maine
3. Ashin, Mark*
University of Chicago
Chicago, Illinois
4. Barry, James D.*
Loyola University
Chicago, Illinois
5. Bliss, Frank W.
St. Olaf College
Northfield, Minnesota
6. Borden, Arthur, R., Jr., Chief Reader
Washington and Lee University
Lexington, Virginia
7. Burbank, Rex*
San Jose State College
San Jose, California
8. Campbell, Jackson J. **
Princeton, University
Princeton, New Jersey

*The readers' names that are starred helped in this study.

**The readers' names that are double starred withdrew as rankers.

9. Coles, William E., Jr.**
Amherst College
Amherst, Massachusetts
10. Coolidge, John S.*
Swarthmore College
Swarthmore, Pennsylvania
11. Goldstein, Malcolm
Stanford University
Stanford, California
12. Hamilton, W. Paul
Montclair State College
Upper Montclair, New Jersey
13. McKenzie, Dorothy C.*
Los Angeles State College
Pasadena, California
14. Main, Charles F.
Rutgers University
New Brunswick, New Jersey
15. Muirden, Marjorie**
Portland State College
Portland, Oregon
16. Roellinger, Francis X.*
Oberlin College
Oberlin, Ohio
17. Royer, R. Christine
Columbia University
New York, New York
18. Schweik, Robert C.
Marquette University
Milwaukee, Wisconsin

*The readers' names that are starred helped in this study.

**The readers' names that are double starred withdrew as rankers.

19. Soehren, Helen*
University of Oregon
Eugene, Oregon
20. Wallace, Robert A.*
Bryn Mawr College
Bryn Mawr, Pennsylvania
21. Whitman, Robert F.
University of Pittsburgh
Pittsburgh, Pennsylvania

The following college English experts were selected from Montana colleges. These experts helped in the study.

1. Keller, Joseph L.
Northern Montana College
Hayre, Montana
2. Noyes, Robert H.
Eastern Montana College of Education
Billings, Montana

*The readers' names that are starred helped in this study.

Letter Sent to English Composition Readers of the
Educational Testing Service

March 30, 1962

(Inside address)

(Salutation)

Since you have been a reader for the Educational Testing Service, I feel that you are especially well qualified to help me with a special project I am doing for my Ed.D. degree at Montana State College.

I have prepared a list of scale points from various literature on how to grade written composition for high school students. This list of scale points is presented on the following page and needs to be ranked in order of importance.

I need your cooperation and will appreciate any help you can give me.

Sincerely yours,

Donald R. Fostvedt

Letter Sent to College English Experts in Montana

March 30, 1962

(Inside address)

(Salutation)

As part of the thesis for my Ed.D. degree, it is necessary that the five items listed on the following page are ranked in order of their importance by several English instructors from several colleges.

I will appreciate it very much if you will be one of the instructors to do the ranking and list the items as you believe they are in importance.

Thank you for your time and consideration.

Sincerely yours,

Donald R. Fostvedt

Ranking Sheet Sent to College English Experts

Please rank in order the importance of the major items listed (number 1 as the most important and number 5 as the least important) and do not mark any that you think should have not been included.

Coherence and Logic _____

Development of Ideas _____

Diction _____

Emphasis _____

Organization Through Sentence
Structure and Paragraphing _____

It is realized that this scale is not absolute, and maybe you feel that one or more items should be differently arranged. If this is so, please indicate this below. Also, indicate any items you feel should have been on the scale that were not there--major items as well as those included below the major headings.

Explanation of Grading Sheet Scale Points
Sent to College Experts

Grading Sheet--Scale Points

1. Coherence and Logic

This includes thought and word coherence, organization of the whole theme (planned logically), and discussion of the subject intelligently.

Coherence is defined as connection or congruity arising from some common principle or relationship; consistency.

Logic is defined as connection, as of facts or events, in a rational way.

3. Diction

This includes use of idioms, word order, vocabulary, and word choice.

Diction is defined as choice of words to express ideas; mode of expression in language.

2. Development of Ideas

This includes overall effectiveness, content (Is the conception clear, accurate, and complete? Is there a significant central idea?), introduction and conclusion, evidence in support of generalizations, originality of treatment, honesty in expression, and sincerity in expression.

4. Emphasis

Has the student used the correct emphasis to make his writing appropriate; to the purpose, and the occasion?

5. Organization Through Sentence Structure and Paragraphing

Marking of Marjorie Muriden, Portland State College, Portland, Oregon,
April 13, 1962, on the "Explanation of Grading Scale Points"

Grading Sheet--Scale Points

1. Coherence and Logic

This includes thought and word coherence, organization of the whole theme (planned logically), and discussion of the subject intelligently.

Coherence is defined as connection or congruity arising from some common principle or relationship; consistency.

Logic is defined as connection, as of facts or events, in a rational way

2. Development of Ideas

This includes overall (effectiveness, content (Is the concept clear, accurate, and complete? Is there a significant central idea?), introduction and conclusion, evidence, in support of generalizations, originality of treatment, honesty in expression, and sincerity in expression.

3. Diction

This includes use of idioms, word order, vocabulary, and word choice.

Diction is defined as choice of words to express ideas; mode of expression in language.

4. Emphasis

Has the student used the correct emphasis to make his writing appropriate; to the purpose, and the occasion?

5. Organization Through Sentence Structure and Paragraphing

"This seems to include all the others."

Educational Testing Service Norms Sent to the Writer by Helen Soehren,
University of Oregon, Eugene, Oregon, April, 1962

Definition of Norms:

1. STYLE: Advanced Placement writing is characterized by precision, economy, interest and vigor in the choice of words. (Verbiage, such as redundancy, academic jargon, and "fine" or "literary" writing, are to be penalized.) While employing a variety of sentence structures, the candidate must be able to control the more complex forms without loss of clearness. The phrasing should be idiomatic; it should avoid faulty and awkward constructions, should have a pleasing cadence, and should be consistent in tone.
2. CONTENT: The question calls for selection of an illusion (or idealization of life) with which the candidate is familiar, and for description of that illusion and of its special appeal. The first requirement is therefore, the appropriateness of the particular subject chosen: does it lend itself to the required treatment? The inherent interest of the material and the resourcefulness with which it is treated are important. Keeping in mind that the amount of detail needed may vary according to the subject chosen, is the detail full enough to give the essay weight?
3. ORGANIZATION: The question calls for a basic two-part or three-part structure of description, reactions and conclusion. Any scheme which embodies this requirement, with whatever variations, is acceptable. A clear sense of progression by well-developed paragraphs is essential.
4. REASONING: The question calls for a definite conclusion. How valid and significant is the candidate's conclusion? How incisive is his analysis of the particular illusion?

The parts of the essay should be logically consistent, and the conclusion should be convincingly established by the evidence provided.

Appendix C

Materials Directed to High School Experts
Pertinent to the Theme Judging

High School English Teachers Who Agreed to
Participate in the Study

1. Mrs. Margaret Maness,* Troy, Montana
2. Sister Teresina,* Havre, Montana
3. Mrs. Ellen B. Skones,* Drummond, Montana
4. Mrs. Helen Smith,* Forsyth, Montana
5. Mrs. Almeda E. Mann,* Columbus, Montana
6. Mrs. Fern Flanagan,* Stanford, Montana
7. Patricia Lee Stevens,* Cascade, Montana
8. Miss Helen Riley,* Butte, Montana
9. Mrs. H. H. Francis,* Arlee, Montana
10. Mr. Robert Hammer,* Sidney, Montana
11. Mrs. Robert Hutchin, Charlo, Montana
12. Mr. Harold Boe,* Miles City, Montana
13. Mr. Ray Mars,* Billings, Montana
14. Mrs. Mary K. Holthaus,* Billings, Montana
15. Mrs. Olive Wehr,* Wolf Point, Montana
16. Mrs. Agatha Dwyer,* Wolf Point, Montana
17. Mr. James G. Sloan, Whitefish, Montana
18. Mrs. Dorothy MacDonald,* Whitefish, Montana
19. Mr. Charles E. Dillon, Victor, Montana

*The teachers names that actually participated in the study are starred.

20. Miss Alicia Bingham,* Victor, Montana
21. Mr. S. Bewley,* Missoula, Montana
22. Mrs. E. Hallford,* Missoula, Montana
23. Miss J. Isobel Eaton, Deer Lodge, Montana
24. Miss Thelma Shaw,* Deer Lodge, Montana
25. Miss Winifred Lapp,* Deer Lodge, Montana
26. Mrs. V. K. Moen, Corvallis, Montana
27. Mrs. Helen Micka, Corvallis, Montana
28. Mr. Robert Olson, Corvallis, Montana
29. Mr. Wilfred Lundberg, Dodson, Montana
30. Mr. William B. Black, Dodson, Montana
31. Sister May Eileen,* Helena, Montana
32. W. E. Lumpkin, Flaxville, Montana
33. Mr. Terence McCourt,* Neihart, Montana
34. Mr. Jim Cox,* Ryegate, Montana
35. Mr. Rudolph Kruger,* Oilmont, Montana
36. Mr. Neil McFadgen,* Havre, Montana
37. Mr. Duane Hoynes,* Fort Benton, Montana
38. Mr. Morgan J. Sherlock,* Conrad, Montana
39. Miss Gertrude Conwell,* Big Timber, Montana
40. Mr. Jack Gardner,* Broadview, Montana

*The teachers names that actually participated in the study are starred.

Instructions to the Supervising Teachers of the Four High Schools
That Supplied Themes for This Study

1. Subject (Neighbors, good and bad) (people)
2. Have themes written between February 26 and March 2, 1962
3. The students will have enough time to complete the themes in not more than one class period
4. Write in ink
5. Length (200-300 words--a page to a page and one-half)
6. All students use the same size paper
7. Have the students put their names on the papers (their names will not be used later except to give their sex in the writing of thesis)
8. Ask students to write as accurately as they know how
9. Do not let students know that this is part of a study--let them think it is part of their regular class work
10. Mail themes right away to Donald R. Fostvedt, Education Department, Montana State College, Bozeman, Montana

Letter Sent to the Heads of English Departments in Montana High Schools
Requesting English Teachers to Participate in This Study

Date: February 26, 1962

To: English Departments
Montana High Schools

From: Donald R. Fostvedt
Education Department
Montana State College
Bozeman, Montana

I am conducting a study at Montana State College to find the significance of a scale for grading the quality of English composition in high schools.

Compositions have been obtained from four Montana high schools from which twenty compositions shall be selected for evaluation.

If you have teachers on your staff that are qualified to participate in this study, their cooperation shall be very much appreciated. The qualifications are as follows: a college degree, a teaching major in English, and at least two full years experience as an English teacher.

Those that take part in the study shall receive the twenty themes about April 1, 1962, and I would like to have them returned by April 15, 1962.

Please let me know at your earliest convenience by return mail if you, or any of the English teachers in your school, will cooperate in this study.

Thank you.

Indicate below the names and addresses of English teachers that will cooperate in this study.

Packet Materials Sent to Individual High School English Teachers
Who Participated in the Study

Education Department
Montana State College
Bozeman, Montana
March 26, 1962

Dear English Teacher:

I wish to thank each of you that agreed to participate in this study. It is indeed wonderful to know that so many of the high school English teachers of Montana are interested in improvement of grading techniques for written composition.

You are asked to evaluate the themes according to scale points, rank the importance of the major scale points, and give each theme an overall rating.

The scale points on the correction sheets have been compiled from criteria established by experts in the English field.

The twenty themes that are enclosed are a random sampling of twelfth grade student's themes. These twenty themes were chosen from about two-hundred and sixty themes submitted by four Montana high schools.

A few of the more pronounced errors in some of the themes were corrected when typing for mimeographing, because we wish to grade for quality of content, not mechanics as such.

I shall appreciate it very much if you will have the grading sheets returned to me by April 15. You need not return the themes, but I would appreciate it if you don't destroy them until after you receive the results of this study.

Thank you once more for your cooperation.

Sincerely yours,

Donald R. Fostvedt

Grading Sheet

Please grade each of the twenty themes according to the five criteria given:

Criterion Number	Theme Number	1. Poor	2. Average	3. Better Than Average
1. Coherence and Logic This includes thought and word coherence, organization of the whole theme (planned logically), and discussion of the subject intelligently. Coherence is defined as connection or congruity arising from some common principle or relationship; consistency. Logic is defined as connection, as of facts or events, in a rational way.	1.	_____	_____	_____
	2.	_____	_____	_____
	3.	_____	_____	_____
	4.	_____	_____	_____
	5.	_____	_____	_____
	6.	_____	_____	_____
	7.	_____	_____	_____
	8.	_____	_____	_____
	9.	_____	_____	_____
	10.	_____	_____	_____
	11.	_____	_____	_____
	12.	_____	_____	_____
	13.	_____	_____	_____
	14.	_____	_____	_____
	15.	_____	_____	_____
	16.	_____	_____	_____
	17.	_____	_____	_____
	18.	_____	_____	_____
	19.	_____	_____	_____
	20.	_____	_____	_____

Criterion Number

Theme Number	1. Poor	2. Average	3. Better Than Average
-----------------	------------	---------------	---------------------------------

2. Development of Ideas

This includes overall effectiveness, content (Is the conception clear, accurate, and complete? Is there a significant central idea?), introduction and conclusion, evidence in support of generalizations, originality of treatment, honesty in expression, and sincerity in expression

1.			
2.			
3.			
4.			
5.			
6.			
7.			
8.			
9.			
10.			
11.			
12.			
13.			
14.			
15.			
16.			
17.			
18.			
19.			
20.			

Criterion Number	Theme Number	1. Poor	2. Average	3. Better Than Average
3. Diction This includes use of idioms, word order, vocabulary, and word choice. Diction is defined as choice of words to express ideas; mode of expression in language.	1.			
	2.			
	3.			
	4.			
	5.			
	6.			
	7.			
	8.			
	9.			
	10.			
	11.			
	12.			
	13.			
	14.			
	15.			
	16.			
	17.			
	18.			
	19.			
	20.			

Criterion Number

Theme Number	1. Poor	2. Average	3. Better Than Average
-----------------	------------	---------------	---------------------------------

4. Emphasis

Has the student used the correct emphasis to make his writing appropriate; to the purpose, and the occasion?

1.	_____	_____	_____
2.	_____	_____	_____
3.	_____	_____	_____
4.	_____	_____	_____
5.	_____	_____	_____
6.	_____	_____	_____
7.	_____	_____	_____
8.	_____	_____	_____
9.	_____	_____	_____
10.	_____	_____	_____
11.	_____	_____	_____
12.	_____	_____	_____
13.	_____	_____	_____
14.	_____	_____	_____
15.	_____	_____	_____
16.	_____	_____	_____
17.	_____	_____	_____
18.	_____	_____	_____
19.	_____	_____	_____
20.	_____	_____	_____

Criterion Number

Theme Number	1. Poor	2. Average	3. Better Than Average
-----------------	------------	---------------	---------------------------------

5. Organization Through Sentence
Structure and Paragraphing

1.	_____	_____	_____
2.	_____	_____	_____
3.	_____	_____	_____
4.	_____	_____	_____
5.	_____	_____	_____
6.	_____	_____	_____
7.	_____	_____	_____
8.	_____	_____	_____
9.	_____	_____	_____
10.	_____	_____	_____
11.	_____	_____	_____
12.	_____	_____	_____
13.	_____	_____	_____
14.	_____	_____	_____
15.	_____	_____	_____
16.	_____	_____	_____
17.	_____	_____	_____
18.	_____	_____	_____
19.	_____	_____	_____
20.	_____	_____	_____

Ranking Sheet

Please rank in order the importance of the major items listed (number 1 as the most important and number 5 as the least important) and do not mark any that you think should not have been included.

Coherence and Logic _____

Development of Ideas _____

Diction _____

Emphasis _____

Organization Through Sentence
Structure and Paragraphing _____

It is realized that this scale is not absolute, and maybe you feel that one or more items should be differently arranged. If this is so, please indicate this below. Also, indicate any items you feel should have been on the scale that were not there--major items as well as those included below the major headings.

Give each theme an overall grade of 1, 2, 3, 4, or 5 with a (1) as the lowest grade and a (5) as the highest possible grade.

Theme Number	1.	2.	3.	4.	5.
1.	_____	_____	_____	_____	_____
2.	_____	_____	_____	_____	_____
3.	_____	_____	_____	_____	_____
4.	_____	_____	_____	_____	_____
5.	_____	_____	_____	_____	_____
6.	_____	_____	_____	_____	_____
7.	_____	_____	_____	_____	_____
8.	_____	_____	_____	_____	_____
9.	_____	_____	_____	_____	_____
10.	_____	_____	_____	_____	_____
11.	_____	_____	_____	_____	_____
12.	_____	_____	_____	_____	_____
13.	_____	_____	_____	_____	_____
14.	_____	_____	_____	_____	_____
15.	_____	_____	_____	_____	_____
16.	_____	_____	_____	_____	_____
17.	_____	_____	_____	_____	_____
18.	_____	_____	_____	_____	_____
19.	_____	_____	_____	_____	_____
20.	_____	_____	_____	_____	_____

Theme #1

Neighbors, Good and Bad

Who and what are neighbors? Neighbors are termed as immediate people living next door. In case of trouble or an accident they are people you can usually depend upon.

To always be next to you when you need a friend the most is a good neighbor, yet he does not have his nose in your affairs. He will lend you instruments or goods when necessary without henpecking you to return them, but he will occasionally let you know you still have them. To depend on at all times is a role often played by a good neighbor. Someone to swap receipts and ideas with over a cup of coffee is a role of a good neighbor. A good neighbor is a friend indeed!

Bad neighbors are your friends, the ones that are so friendly they know all your private affairs, and see to it that everyone does, too. He's a person that loves to lend you needs, but boy, how he lets you know it, henpecking he tells you frequently that you owe him this, you borrowed this, and never ceases to annoy you with his deeds. A perfect neighbor and citizen, the bad neighbor tells the truth at all times, well almost all the time, just a little exaggeration here and there. Lovingly he pats you on the back with his right hand, while he steals you blind with his left. Yet, this bad neighbor will help you to his full ability whenever help is needed.

Neighbors good or bad, are something that this modern day society

Theme #1 (continued)

needs and depends on. Loving, crafty, and prejudice describes our neighbors, yours and mine, not just local, but national and world wide, for today's society depends on all nations with its modern conveniences.

Theme #2

Neighbors, Good and Bad

The neighbors that I have are mostly real nice people. The neighbors on our right have two children. The oldest child is five and has been blind all her life. The youngest child is two. They also have a white German-shepherd dog, who always follows me around. The people on our left are new. They used to live up around Kalispell, but moved down here when they bought a store. They have three children and the kids are well behaved most of the time.

The people who live in back of us are sometimes very noisy. The father and mother drink quite a bit and fight even more. They have rabbits and dogs and neither is penned up. They also have about six kids--two boys, who are devils most of the time, and four girls. The neighbors in front of us live in apartments and work or go to college, and I very seldom see them.

Our one block has about forty some children on it and twice that many dogs. I think neighbors are essential, but not really necessary. They can give some people a sense of security, but to me, most of them give me a headache. Just about all the kids on my block are real young. I have to go three blocks both up and down before I meet someone my own age. My family has lived in the same neighborhood for almost seven years now. All the grown-ups around us work--including the mothers.

I think to be a good neighbor you have to be considerate of the

Theme #2 (continued)

ones living around you. You must not think entirely of your own needs, but those of others too.

Some of our neighbors don't--meaning the ones who live behind us. They burn awful smelling garbage. Their yard looks like white-trash live there and they swear so loud that the whole block can hear them.

I enjoy all the rest of my neighbors, but I don't make a habit of going visiting and vice-versa. I also do very little baby-sitting, because they know I have plenty to do with keeping my house and brothers and sister up, and also going to school and keeping a job.

I would say on the whole--my neighbors are real swell people.

Theme #3

Neighbors, Good and Bad

Neighbors are usually pretty good people to have around. If one runs out of people to talk about there is always something to say about a neighbor. When there is nothing to do at home you can always go to one of the neighbor's places and bother or help him. When a person is in trouble he can usually count on his neighbors.

In 1945 our house burned down, and times were pretty tough, but the neighbors took care of us for about two weeks--and at that time that's about all they could do is to feed themselves. They shared their food and home willingly.

With neighbors there are usually kids. They are always more fun to play with than your own brothers or sisters. But, as soon as you do something wrong they tell their parents and their parents tell yours--and then you're in trouble. There is no comparison between the neighbor of a large city and a small town. In the large city nobody knows his neighbor and most of them don't care to know him. In a small town everybody knows everybody and what they do all day long. That is why there is so much gossip in a small town.

Theme #4.

Neighbors, Good and Bad

What are neighbors? Well, neighbors have a different purpose and meaning depending on where you live or how you use the word.

If you live in a city, your neighbors live in the house next to yours or in the same block. Your neighbor's dog may bark all night and their children may throw footballs through your windows. The lady next door may take singing lessons and practice all day long on her scales and exercises. But on the other hand, whenever you need help or advice, your neighbors are the first to the rescue.

If you happen to live on a farm, neighbors mean something quite different. Your neighbors in the country may live ten to twenty or five miles from you. You see them only very seldom and visiting your neighbors in the country is a big occasion. These neighbor's cattle may break down a fence on your property and feed on and trample down your crop. They may borrow a piece of machinery and never bring it back. But neighbors in the country will do your chores while you're on vacation, help you finish your harvesting and lend you the shirt off their backs.

Neighbors to me mean all these things, good and bad, but the good always overrides the bad.

Theme #5

Neighbors, Good and Bad

Nearly all of us have neighbors of some kind. Our characteristics are the main factors which determine whether we are "good" or "bad" neighbors.

Our natures determine who we think are our neighbors. To some people everyone in the world is a neighbor. Other people set up arbitrary geographical limits--for example, everybody living within a ten-mile radius of a person's house are his neighbors. Still others limit their neighbors to a certain race or religion. Some do not regard anybody as a neighbor.

Those in the last category usually make the poorest neighbors. They are unfriendly; they are constantly bickering with someone. Some of these people continually stir up commotion with no regard for the peace of the neighborhood. They tend to make life unpleasant for those near them.

Those who are prejudiced against a certain type are usually inconsistent and hypocritical, for they are pleasant and sweet to their own type, but they are obnoxious to other types.

On the other hand, the unprejudiced, friendly people who consider the whole world their neighbor tend to make life more pleasant for others. They enjoy being with people and spread cheer and warmth wherever they go.

Thus, the more friends people think they have, the more they do have.

Theme #6

Neighbors, Good and Bad

I feel that a warm, friendly relationship with the people living about you is very important. In order to build up these relationships there must be cooperation, understanding and the ability to compromise.

It is always said that every family should have children and these children should have pets. Many disagreements between neighbors stem from either of these two sources. This situation occurs when homes are too close together for anyone's comfort. To settle these types of problems parents and friends should look at both sides of the problem and try to settle it reasonably.

It is often said that no one can live without friends. A person who can live in a community and not associate with his fellow neighbors will certainly be unhappy. If a person realizes or finds he cannot get along with the people in his neighborhood he should try to find the source of trouble.

How can a person determine whether he is a good neighbor? It is an old saying that money cannot buy love and friendship. You must achieve these things by giving yourself, not your money. If a person has a pleasing and warm personality he can be considered a good neighbor and an asset to his community.

Theme #7

Neighbors, Good and Bad

I have one certain neighbor whom I absolutely hate. She is about ninety years old, and she is about the crabbiest thing on earth. If you get within thirty feet of her house she comes barging out the door to see who you are. She is almost totally deaf. If she has her hearing aid turned up high and if you yell as loud as you can into her ear, she can almost understand what you are saying.

This woman, of course, is one of those old bitties who thinks that the town owes her everything. She is never satisfied with the way the town is being run.

On the other hand, I have a neighbor who is the nicest man I can think of. He never gripes about anything, and he is content to let the proper authorities run the town. He is old too, about ninety, but he isn't crabby or suspicious.

I think that I have two perfect examples of neighbors of being on the two extremes. One is the most crabby person living, the other the nicest. I think that this is true in the United States as a whole because of different people being taught different things.

The two neighbors that I have are sort of different though, because they are married to each other.

Theme #8

Neighbors, Good and Bad

To one side of us lives a young divorcee with a half dozen brats. She doesn't make them mind her or anybody else either.

The kids walk out or run out into the middle of the street just asking to be hit, and usually during the most busy parts of the day when traffic is the heaviest.

They, the children, pick fights with all the other children in the neighborhood. When the kids ask them to join them in a game they don't agree on how the game should be played; they say it is a sissy game, or they always want to be the person who leads the game more or less, or the person who is it.

Then if they don't feel like exercising very much they get rocks from the alley and throw them at the cars as they go by.

It seems to me that the mother should be able to make them mind better than that. And she doesn't work so she is there most of the time to watch them.

On the other side of us is an old couple who keep pretty much to themselves. They are very friendly and nice to talk to. But they aren't trying to snoop into anybody's business all the time like some people do.

Then in back of us there is this old couple who are very friendly. They are always ready to help if the need arises. They are the kind of people that you can talk to over the fence on a nice summer day. They

Theme #8 (continued)

are agreeable and quick to see your point of view, but they expect you to see theirs too. If they don't agree. This is only right though.

Well, what can you say about neighbors? I don't think that there is an ideal neighbor, and it would probably get mighty boring if all your neighbors were ideal. Your neighbors are what they are good or bad, and you have to make the best of it.

If you complain and fight with them all the time it affects everybody. If the parents fight all the time they make their children stay away from each other. So don't fight with your neighbors, except for your rights. It isn't worth it and the rest of your family just gets the worst from it.

Theme #9

Neighbors, Good and Bad

In the Bible one is taught to love his neighbor as himself. One then is caused to wonder who is his neighbor. This wasn't explained in terms of good or bad neighbors but explained in the Bible with whoever is in need of ones neighbor. Not just the person across from one or the person he lives next door to, ones neighbor may be the starving family at the other end of town or other side of the world. One can't decide whether his neighbors are good or bad because of race, color, or creed. He should love them all.

One in school has so many neighbors because he attends every class with a different neighbor. They can be good or bad depending upon his taste of friends.

If one happens to be alone in a strange country he may turn to his neighbor for aid. His neighbor may speak a strange tongue, but he is still a neighbor. The neighbor may be helpful to one and turn down help to another, because he wasn't polite asking for aid. One cannot say he is a bad neighbor because he helped one and not another.

One must be a neighbor to have a neighbor so whether it's the person across from one, next door, or across the world if he wants a neighbor--good or bad--he'll have to be a neighbor first and remember to love his neighbor as himself.

Theme #10

Neighbors, Good and Bad

A neighbor is one who, in general, will mow the lawn while you're on vacation, and take care of your trained dog. He, for the most part, is a likeable character; he knows when to mind his own business and, of course, knows when he's not welcome. He is active in the local activities and almost always never misses a P.T.A. meeting. He is a firm believer in establishing good relationships with his children.

The ardent church-goer is he; never misses a Sunday unless it's for one of those early fishing trips with the boys. He'll cover-up for you if you've had a fight with your spouse, but he doesn't like to get tangled up in a family war. To size him up very quickly, we could say he's an all-around good Joe.

The bad neighbors are easily spotted in a neighborhood. He's a pill and his wife is a talking tape recorder. She never shuts up for a minute and is constantly on the telephone. He can't stand to have his flower-beds disturbed, so what does he do? He sends his little bundle of joys over to play in your yard. He has an over-sized dog, a mongrel as a matter of fact, that loves to raid your garbage can. And he's quite clever at this for it's not every dog that can scatter your garbage all over the yard. How clever can you get!!

His wife is one of those women who is a window-watcher; she knows what you're doing every minute of the day, even when you use the toilet.

Theme #10 (continued)

Your neighbor, all sized up in one package, is a big bore with a wife who is too terrible for words.

Theme #11

Neighbors, Good and Bad

A neighbor is a person with whom you associate day by day. We associate with good neighbors and bad neighbors. We should know how to get along with each kind. To do this I feel we should know what good neighbors and bad neighbors are--in our own minds.

I feel a good neighbor is a person with whom you can agree and come to a mutual understanding without hurting each other. I don't mean by hurt, physical blows, but things said or done.

A good neighbor is one that respects your beliefs as to religion, morals, ideals, and so on. If you are of two different religions he won't force his on you or he won't try to change your goals in any way. He should be treated by you as a brother and you should be able to confide in him, and when in need be able to go to him for help. But to be a good neighbor yourself you shouldn't burden your neighbor with every little problem that comes up.

A bad neighbor is a person that when you confide in him he discusses your problem around with some of his friends with whom it doesn't concern.

He may be the type of person that is always gossiping or telling you all of his troubles. He does things on purpose that interfere with your goals or line of reasoning. To be brief, he is more bother than he is worth.

Theme #11 (continued)

Good neighbors can help each other quite a bit, but one selfish or poor neighbor can ruin everything and what they stand for.

Theme #12

Neighbors, Good and Bad

Neighbors are persons who live near each other, and many of their activities are based upon one another. One can be blessed with good neighbors, or one can be afflicted with bad neighbors.

Good neighbors are enjoyable, and the family can have many good times together. Often people like to spend their leisure time with their neighbors, enjoying some mutual activity. Also, neighbors can really be friends in need. They can help each other with their work and get things done faster and have fun doing it.

Bad neighbors can ruin a family's fun and make life miserable for everyone. There are many ways in which bad neighbors can contribute to the delinquency of the children. They can do this by setting poor examples for children, and children are easily susceptible to the traits of older persons. Bad neighbors will not bring up their own children properly, and thus some of this will rub off on other children. Bad neighbors will not help someone in distress as will good neighbors; they are concerned primarily with themselves. In a city bad neighbors do not contribute to the well-being of their neighborhood. They do not try to keep their homes and yards in good condition and are not concerned with a better life for the whole community.

Good neighbors wish to promote their neighborhood and keep it in top shape. They like to help their neighbors and feel that they should

Theme #12 (continued)

receive this same treatment in return.

Being a good neighbor should not be limited to the boundaries of our country, but it should be an important policy that we extend to all other peoples.

Theme #13

Neighbors, Good and Bad

Neighbors are not only persons who live next door to you. In my opinion good neighbors can live clear across town. Bad neighbors can live only on your block in town.

Having good neighbors means to have good friends and be in friendly fellowship with others. Good neighbors are not for keeping the next house clean or to make yours nice with a cut lawn, but they are for understanding and respect and ideas. Neighbors are not only for the extra cup of sugar you need, nor a baby sitter in an emergency, but they are a necessary organ of the community!

Neighbors are human beings with human faults. Naturally many people have more than their share of faults. If, by unfortunate chance, one lives beside you he is classified as a bad neighbor or just a plain grouch. The reasons for this bad neighbor may not be clear to his other neighbors, but to him they may be justifiable--in his eyes. To him the community may be a despicable, annoying nuisance which he wants nothing to do with. In his eyes babies crying, kids ruining his lawn, dogs howling, and teenagers are his mortal enemies. He wants a secluded, quiet life which seems strange to us. His private life is being invaded by these children, dogs, and teenagers, so he becomes "grouchy".

Neighbors, good or bad, belong in a community or any society. Bad neighbors may be helped by becoming good neighbors yourself.

Theme #14

Neighbors, Good and Bad

We all have neighbors here in our town. We exchange friendly talk about subjects of common interest while we work together to get a stubborn lawn-mower working, or while we participate in a friendly game of cards. Our good neighbor sees us walking on a cold, rainy day; he stops and offers us a ride. He smiles as he passes us on the street; our whole day seems to brighten and gain a new perspective. He is always ready to give a helping hand when it's needed; he's a friend.

But, a neighbor is more than the fellow who happens to live next door or across the street. We have never seen some of our best neighbors and will never know their names. Our good neighbors are those men everywhere who, no matter against what odds, still retain their own thoughts and strive for a better life.

Our good neighbor may be the hungry Chinese farmer who toils ceaselessly against cruel task-masters, nature and man, and yet, in his own way, and in the recesses of his heart, sends up a hopeful prayer that all men might be free from subjugation.

He might be one of the cocoa-brown peons, working hard in the sweating tropics so that people he has never seen might enjoy his bananas, or coffee. He wishes that people everywhere might feel as happy and care-free as does he.

Our good neighbors whom we have never seen wish for us the same as

Theme #14 (continued)

we do for them--that men might someday live together peacefully and help-
fully.

Theme #15

Neighbors, Good and Bad

Although physical features of an area play a large part in the area, I believe that your neighbors have a large role concerning your happiness in a neighborhood. It's probably to assume that the more you stay at home the more contact you will have with your neighbors.

A neighborhood of warm, friendly people can indeed be a blessing. Understanding and group cooperation can lead to a full, satisfying group relationship. To achieve this relationship a few conditions should be met. Neighbors that have many things in common usually get along a lot better than others. A family with many young children may not fit in an area of retired people. An older couple may not appreciate all the neighborhood children.

Good neighbors many times have fun together. A neighborhood barbeque or an evening get-together can add much to the enjoyment of group living. Good neighbors usually are responsible neighbors. Late, loud parties in an area where many people go to bed early are not really appreciated. Good neighbors usually take pride in the neighborhood. They try to keep the neighborhood attractive by keeping their yards in a pleasing state. A neighbor whose yard is messy and unkept doesn't give much to the appearance of yours.

Being a good neighbor is something every family should try to be. A neighborhood of happy "sugar-borrowing" families can be a wonderful

Theme #15 (continued)

place in which to live.

Theme #16

Neighbors, Good and Bad

A couple of years ago my grandfather had a stroke. Being at Christmas time, Mom and Dad couldn't be home. Having heard about this mishap, our neighbors offered to cook our dinner and get everything ready for Christmas day.

Our neighbor's little girl was having her tonsils out and couldn't be left alone. To sit by her daughter with nothing to eat was a tedious job. My mother offered to sit with her daughter a couple of hours a day, so to give her time to eat, etc. Our neighborhood gets along fairly well, considering all the little kids around.

In a neighbor's yard which was forbidden to go into, the children of the neighborhood decided to play. Seeing all the flowers she had planted, the children picked them to bring to their mothers. When she saw the flowers had been picked, she went to each mother and collected five dollars from each. She must have very beautiful flowers for the price she charged for the picked ones.

All in all, we have to have both good and bad neighbors to balance out the neighborhood so that it doesn't get boring.

Theme #17

Neighbors, Good and Bad

When we lived on Second Avenue, the neighbor on our right was one of the best. While the neighbor on our left was the worst.

Harry, the neighbor on our right, was an old widower and the kids, it seemed, could do almost no wrong. He took me fishing a couple of times and though we didn't catch much, I learned a lot. When the time for crabapples came we all went to see Harry. He said he didn't mind as long as we were careful not to break any limbs off the tree. During the summer Harry liked to tell stories, so a little after supper most of the kids would be over there. Everybody, as far as I know, liked Harry.

The Smiths lived on our left. They were an old couple that had retired from the farm and moved into town. They were always griping about us kids making too much noise. On the left side of our house was a wide and long stretch of lawn, and we liked to play football there. Well, we'd be playing football and anywhere from one to two minutes after we would have gotten started Mr. Smith would come out and tell us to 'git'. My folks owned that stretch of lawn, but when he came out we would leave. One day when he came and told us to 'git' we decided to stay and play. Well, he came out about four times before he called the cops. When the cops came we kept on playing. The football got loose and rolled onto Mr. Smith's land. I took two jumps, picked up the ball, and jumped back. The game continued. During all this time the police had been sitting in

Theme #17. (continued)

their car at the corner. Mr. Smith came out, looked at the police, then at us, then the police. The police then drove away and Mr. Smith went back into the house, very mad. We kids realized that we could do what we wanted as long as we stayed on our own property, and so we made life very miserable for Mr. Smith.

Theme #18

Neighbors, Good and Bad

The neighborhood I live in is like "one big happy family," each of us being a brother or sister to the others. Doing helpful things for each other, talking problems over, drinking coffee together and discussing everything from world problems to domestic problems are characteristic of "the block". During the summers barbecues are held for no special reason except for a get-together. Every family brings something and in no time everyone is having a good time.

Helping with the children, cooking the meals, and washing clothes are just a few examples the neighbors share when a family has had bad luck or an illness. To take in a wash that is hanging on the lines or to close any windows that might have been left open when a storm unexpectedly comes up means very much to that neighborhood. An act of this nature has happened very often in the neighborhood.

Because there are close to thirty-five children on the block, everyone has grown accustomed to the noise. Hearing them at play reminds me of recess time at school. If the neighborhood was quiet, it would be very odd.

One family complained of the noise. Mr. "X's" children weren't allowed out of the fence and suffered very much. The parents never did understand that the other "monsters" on our block were good and just being themselves. I think it's too bad when parents won't permit their

Theme #18 (continued)

children to run with the others as they did. Soon their children became withdrawn.

Except for this incident with family "X", the neighborhood has been peaceful for the twelve years we've lived there. These are neighbors that are really wonderful, and make a family proud to say they live on "the block".

Theme #19

Neighbors, Good and Bad

Neighbors can be very swell people, but they can also be very bad.

Now what would the good neighbor be like? Well, it would be kind of hard to say, because people are not alike in most respects. The good neighbors like to visit and just pass time away--especially in the mornings--say on Saturday morning. Drink coffee and smoke cigarettes, probably your cigarettes, are the two Saturday morning occupations of neighbors. They like to borrow a cup of sugar, flour, etc. But without this kind of person what would you have? -- Probably nothing. Not to have good neighbors is like having nothing to live for. You need friends and friends need you unless they're old crabs or grouches.

The crabs and grouches make it bad for everyone. Now, say you've just moved into a new block or even a new town. Everyone is a stranger. The couple next door are in their mid-forties, a little older than you. Of course you try to make friends by talking over the back fence with something like this. "Hello there, I've just moved here and my name is Van Gorden." By now you have noticed that he was watering the lawn. He doesn't reply, looks at you and kind of makes a face. Then he just keeps right on with what he was doing. Well, you think, this is some person I have for a neighbor. I wonder if he likes people or even himself. He's one of the kind who doesn't like people other than himself. You might call him "take all he can get, but give nothing."

Theme #19 (continued)

We all need others to get peace of mind and the right people will help if they can. This is true in all good neighborhoods and of good neighbors.

Theme #20

Neighbors, Good and Bad

A person can have very nice neighbors, or he can have neighbors that are hard to get along with.

On my block, a new neighbor moved in and at first I thought she was a nice elderly woman. She would come over and visit and have coffee. Sometimes she would play the piano as she said she liked the piano. I never thought too much about it at first, until it got to the point where she would come over any time, no matter what I was doing, or if I had company. I finally got to the point I even hated to see her come. She must have finally realized it, because she started doing anything that would make me angry. First she would holler at my two small brothers that were playing in the yard. When I asked her the reason for this, she said she couldn't rest. I then told my brothers they would have to play quietly. Then one morning I noticed some vegetables were taken from the small garden out in the back of my house. I never said anything at first, because I thought maybe the other children in the block had come over and picked them. But, as things happen, I finally learned the truth. The new neighbor that was so nice at first had stolen them and was still doing it. It wasn't the idea of the vegetables, it was that she took them without asking. I would have given them to her if she would have asked me. I finally caught her at it one day and she never did it again. Then what do you think she did next? She tried to cut the limbs off the trees we had.

Theme #20 (continued)

planted. She said they made noise on the side of her house when the wind blew. We finally got her to realize that the trees would do her more good than harm. She never spoke much after that, as she was very angry. I could go on and on about her, but there is no point in it, because she was just a bad neighbor.

Across the street an elderly couple lived and they were such wonderful neighbors. They often times came over to visit or ask us over to visit them. They would always tell us to come over and watch television. Once in a while they would care for my younger brother while I went shopping on Saturday afternoon. They always were friendly no matter when we saw them or where it happened to be. I guess a person has nice neighbors also.



3 1762 10010419 7

D378

F819

cop.2

Fostvedt, Donald R.

Criteria for the evaluation
of high school English

NAME AND ADDRESS	
6-26-64	Dale A. Tash 321 8007
7-7-	B. Johnson 9 Jeff
Jan. 1965	Fred Leach 2219 West Cr
	Bruce Ayres 1525 E
8-13-64	M. R.
	m. h.
7-2-65	John
10-11-65	

D378
F819
cop.2

APR 1'

I