



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 DeSmet
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

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.

²⁰ ibid., p. 2.

²¹ ibid., 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 brokendown 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

⁴⁹ ibid., p. 13.

⁵⁰ Indiana 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.

