

AN OBJECTIVE SCALE FOR
EVALUATING ENGLISH RESEARCH PAPERS

by 6791

LAWRENCE JAY LEWALLEN

A. B., Washburn University, 1964

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

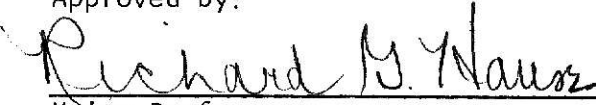
MASTER OF SCIENCE

College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1971

Approved by:


Major Professor

**THIS BOOK
CONTAINS
NUMEROUS PAGES
WITH THE ORIGINAL
PRINTING BEING
SKEWED
DIFFERENTLY FROM
THE TOP OF THE
PAGE TO THE
BOTTOM.**

**THIS IS AS RECEIVED
FROM THE
CUSTOMER.**

L0
2668
T4
1971
L48
C.2

ACKNOWLEDGMENTS

In preparing this thesis I am indebted to many people. The teachers, assisting in the grading, Mrs. Janett O'Hair, Mrs. Eloise Rinne, Mr. Marvin Keith, Mrs. Nancy Keith, and Mrs. Joyce DeBord were most helpful, and the work was greatly appreciated. Thanks also goes to Dr. John T. Roscoe for his assistance and patience in the computer and statistical work for this thesis.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	ii
LIST OF TABLES	iv
LIST OF FIGURES	v
Chapter	Page
I. BACKGROUND FOR THE STUDY	1
Purposes of the Study	2
Definition of Terms	2
II. REVIEW OF RELATED LITERATURE	4
III. METHODS AND MATERIALS	10
Samples	10
Variables	10
Procedures	11
IV. ANALYSIS OF THE DATA	17
V. SUMMARY AND CONCLUSIONS	31
BIBLIOGRAPHY	35
APPENDIXES	38

LIST OF TABLES

Table	Page
1. Descriptive Survey of the Totals - Number Grade	18
2. Descriptive Survey of the Totals - Letter Grade	19
3. Grader Analysis of Mean and Standard Deviation Scores for Objective and Subjective Methods	27
4. Papers Analysis of Mean and Standard Deviation Scores for Objective and Subjective Methods	28
5. Summary Data for T-test for Two Related Samples	29
6. Summary Table for the Analysis of Variance - Objective	29
7. Summary Table for the Analysis of Variance - Subjective	30

LIST OF FIGURES

Figure		Page
1.	Research Paper Evaluation Study Objective	
	Grading Scale	14
2.	Positive Correlation of the Objective and Sub-	
	jective Methods of Grading Research Papers	20
3.	Grader One - Positive Correlation of the Objective	
	and Subjective Methods of Grading Research Papers	21
4.	Grader Two - Positive Correlation of the Objective	
	and Subjective Methods of Grading Research Papers	22
5.	Grader Three - Positive Correlation of the Objective	
	and Subjective Methods of Grading Research Papers	23
6.	Grader Four - Positive Correlation of the Objective	
	and Subjective Methods of Grading Research Papers	24
7.	Grader Five - Positive Correlation of the Objective	
	and Subjective Methods of Grading Research Papers	25
8.	Grader Six - Positive Correlation of the Objective	
	and Subjective Methods of Grading Research Papers	26

CHAPTER I

BACKGROUND FOR THE STUDY

One of the major tasks of the teacher of English is to determine the curricular division of composition, grammar and literature. These divisions are usually based upon the teacher's abilities and preferences. One particular problem encountered in the teaching of composition (and this problem may be the reason many teachers neglect this area) is the grading of these compositions. The teaching and evaluating of composition has received considerable criticism in the past few years. Much of this criticism is justified because the field is almost totally void of any uniform method of grading that helps to eliminate the subjectiveness of grading student work. There has been some research in the field of essay evaluation scales, but very little successful work has been done toward developing a scale for the longer and more involved research writing done by most students in English composition classes.

A scale for the grading of library research papers is needed and, if found more reliable than the presently accepted method of a subjective teacher grade, would produce more reliable grading techniques for teachers of composition.

Purposes of the Study

What difference, if any, will there be in the reliability of grades on student term papers if graded by a subjective method or by an objective scale?

It was the purpose of this study to develop an objective scale for the evaluation of research papers at the high school senior and college freshman level. Research has been conducted in evaluation techniques for compositions, but the longer, more involved, paper lacks acceptable criteria for objective evaluation. These basic considerations were part of the study:

- 1) selection of criteria for a scale
- 2) weighing of the criteria
- 3) development of an objective scale
- 4) establishment of scale validity
- 5) analysis of reliability

Definition of Terms

1. Student. The student referred to in this study was the student in English composition class in grades 12 and 13. The more general use of the term applies to any student taking English composition.

2. Term Paper. The term paper was the longer composition using library sources, footnotes, and bibliography for documentation of the work.

3. Subjective Method. The subjective method was a method of grading term papers lacking specific criteria for assigning grades on the papers.

4. Objective Scale. The objective scale was a scale developed from a valid set of criteria for a more uniform method of evaluating student research papers.

CHAPTER II

REVIEW OF RELATED LITERATURE

In order to produce a more reliable measure of student achievement in the field of English composition, it was necessary for research to be conducted in the area of objective measures of student work. There has been some work done in this area; however, the related literature for this study deals primarily with research in the objective evaluation of short essays and themes. Because of the lack of research in the area of evaluation, an objective rating scale for grading research term papers has not been developed.

The recognition of the problem of grading essays and themes was a slow process and was largely ignored by the English teachers in the past. At the turn of the century, some writers felt the need to develop some form of grading to produce an objective grade on the papers written by students. These studies as reported by Earl Hudelson (8) included several scales and revisions of scales to aid the English teacher in the objective grading process. More recently the research on the College Entrance Examination Board (7) essay section has produced some findings that strengthen the use of objective evaluations.

To date, research has neglected the longer form of written composition and concentrated its effort on the correction of essays and themes. Most of the present studies have been limited to the local school and have not been affected by nation wide development.

This move to individually developed scales has produced several studies. One was the study by D. R. Fostveldt (6) with a group of Montana students and teachers. Fostveldt's consideration was that after validity and reliability were established for a set of criteria for grading student compositions "this criteria could be considered as a standard scale for the evaluation of English composition." (6) The set of criteria as ranked by college experts and high school English experts was:

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

The establishment of reliability of the criteria was done by selecting 20 themes (200-300 words in length) from a larger population. These themes were graded by 30 teacher experts. Fostveldt found that his scale was too loosely developed and was generally not agreeable to the rankers. The findings pointed out that there were differences among the teachers using the scale, and the themes were not equal in quality. The teachers were definitely committed to an objective scale, but despite the effort to develop a scale, reliability of the criteria was not found.

One of the scales used by some English teachers was the California Essay Scale (11) developed by the California Association

of Teachers of English. The scale covered three major areas: 1) content, 2) organization, and 3) style and mechanics. The scale was accompanied by a set of norms consisting of graded essays with comments. The scale did not suggest any sort of weighing of the three areas and treated them as equal.

Another scale was the Cleveland Composition Rating Scale (11) developed by the Cleveland Heights-University Heights City School District. The scale was divided into three areas: 1) content, 2) style, and 3) conventions. The three areas gave percentage weights of 50%, 30%, and 20% in the order of the three areas. The internal divisions used a five point scale to mark the individual items. The internal marking scale was a development made use of by many of the scales. However, the number of internal scale points varied considerably.

The Diederich Rating Scale (10) was also divided into three areas and scores on an internal five point scale. The major areas were: 1) content and organization 50%, 2) style 30%, and 3) mechanics 20%. The scale, after computing the internal scores, used a multiplication factor that converted the scale into a total 100 point scale.

These scales are only examples of contemporary efforts to develop objective scales to aid the teacher in grading compositions.

A recent study by John C. Follman and James A. Anderson (15) compared the reliability of five rating scales. The scales used were:

1) Cleveland Composition Rating Scale, 2) California Essay Scale, 3) Follman English Mechanic Scale, 4) Diederich Rating Scale, and 5) the "Everyman's Scale." The scales had been previously tested for reliability, and this research was checking for the relationship among the five scales. The raters used the scales to develop inter-correlations among the five scales. It was found that all essays had received substantially the same scores from all five scales. There was a high intercorrelation among all of the rankings with the exception of the Diederich Scale. This seems to reaffirm the need for some kind of objective instrument, regardless of the instrument. The research on scales by Follman and Anderson reaffirms the need for an objective instrument for the evaluation of writing. It was found that the reliabilities of the scales differed, but the use of a scale produced more reliable evaluating than did the subjective grade given by most teachers of English composition.

A recent effort to upgrade the evaluating of term papers was made by A. Reed Morrill (12). He created a scale for the objective grading of term papers. The scale was divided into four areas of grading: 1) form, 2) organization, 3) scope and effort, and 4) content. The scale had an internal scaling of six items ranging from zero to twenty-five. The four major areas of grading were equal, and the total point count was 100 points for the entire paper. With the college experts commenting on Fostveldt's (11) scale general agreement showed that the equal weighting of divisions, as in Reed's scale, would not generally be a compatible issue with the teachers of

English composition. Because of the degree of importance teachers place on an item such as Content and Form, the equal weighting Reed used was questionable. He did, however, establish some objectivity to the grading of research papers by using a scale. The individual teacher will have to decide upon the rank and weight of individual items as the assignment and its purposes dictate.

In the numerous discussions as to the need for an objective method of grading compositions, Schumann (14) pointed out the unanimity of grades in a study of grading a composition by several English teachers. The grades of "A" to "D-" were given an essay, later discovered to have been written by a college professor. The essay was not written as a joke but was to show the need for more objective systems and criteria of grading compositions. Further research by Schumann (14) revealed the inequalities of evaluation because of certain non-writing considerations:

- 1) lack of computer assignment of English teacher
- 2) teachers who give no "A" or "F" grades
- 3) repetition of high grades because of student's past record of composition work
- 4) disturbing classroom behavior influences grades on composition papers
- 5) non-writing items, ie., penmanship, neatness, etc.

All of these tended to unduly influence teachers in evaluation of compositions. The factors of name and penmanship tended to influence

the teacher and could be eliminated by typing and numbering to further the objectiveness of grading.

Schumann (14) stressed that grading not be placed upon the whim of the student's teachers. The objectivity of grading the English student must be refined to assure that teachers are as fair as possible in assessing the work of all students. This further exemplified the need for research in the evaluation of compositions.

In summary, the problem of objective grading scales for English composition is still one that is unsolved. Research has been sparse, and the scales developed lack universal acceptance by English teachers. The development of reliable criteria for evaluating library term papers is the least mentioned and researched of all forms of evaluation. Since the term paper is the longest form and is usually graded as more important than other forms of composition, there is a definite need for additional research in the area of composition evaluation.

CHAPTER III

METHODS AND MATERIALS

Samples

The development of an objective scale for the evaluation of term papers was initiated by soliciting responses from professors, associate professors, and assistant professors of English at Kansas State University, Manhattan, Kansas. This group of professors was the sample of experts from the total population of college English teachers. The sample population of college professors is listed on the address sheet (Appendix A).

The selection of the term paper samples was somewhat restrictive because of the availability of a uniformly assigned set of papers. A larger sample would have been desirable. The papers were written by 25 junior college freshmen at Colby Community Junior College, Colby, Kansas. This group of papers was chosen to represent the basic library research paper written by first semester college freshmen.

Variables

The independent variables, as stated in the problem, were the subjective method of grading and the use of an objective scale of criteria for grading research papers. The criteria for study was the grade for both the subjective and objective method, as assigned by the graders.

Procedures

The development of an objective scale was the major task in the initial stages of this research project. Rather than using the opinions and prejudices of the researcher (as the basis for the objective scale) a more valid method of development was necessary. Since the basic consideration was not "the scale" but rather "any objective scale," it was decided to enroll the experience and education of experts in the field of English, ie., professors of English at the university level.

This initial involvement was made of 23 professors of English. The letter of May 26, 1970, was sent to these experts (Appendix B). They were asked to respond by contributing the items they felt were important in the grading of a basic library research paper. The Professor Response Sheet was to provide the returnee with the assignment information designed to assist him in commenting on items for the scale (Appendix C).

The return percentage of the first response was 52%. Major classifications of items were divided into 4 areas: 1) Content, 2) Organization, 3) Form and Style, and 4) Mechanics. From the comments made by the respondees, these major areas were sub-classified, and a preliminary scale was developed.

The Cover Letter, Criteria for Development of the Scale, and Research Paper Evaluation Scale-Percentage Response, were then mailed on July 8, 1970, to the 23 original professors of English (Appendixes D, E, and F).

Of the 23 letters sent 8 were returned, yielding a second response of 35%. The results of this response and the percentage weightings for the four groups were:

I. Content	35%
II. Organization	30%
III. Form and Style	15%
IV. Mechanics	20%

The preceding percentages were the averages from all returned percentages. With this necessary information an objective scale with properly weighted items was constructed.

In the construction of this scale, it was thought that to be effective it must be clear and concise enough for a teacher to feel it worthy of use. Again it needs to be stated that the percentages for each division and the actual completion of the subdivisions was a very arbitrary exercise. For this research it was completed by the college professors, but the task could be done by anyone using the scale. The important consideration in the use of a scale is not that the developed scale be used, but that uniformly developed criteria and the graded worth of these criteria be used. The scale's worth seems to be the uniform method of applying objective grades on a particular item rather than the arbitrary application of some subjective grade on unidentified criteria.

The Objective Grading Scale, Figure 1, was developed into four classifications. The scale uses a 13 point (A+ to F) grade

equivalent scale. Sub-classifications were developed for the four major considerations. The first, Content, was divided into six sub-classifications with a 6% total grade weight for each of the six sub-classifications. Second, Organization, was also divided into six sub-classifications with the percentage at 5 for each item. There were four sub-classifications for Form and Style, with a 4% weighing of the total grade assigned to each of the items. The last classification, Mechanics, had four sub-classifications, and each was worth 4% of the total grade.

This percentage weighing of the sub-classification items did alter the total percentages possible.

CLASSIFICATION	NO. OF SUB-CLASSES	%	TOTAL %
I. Content	6	6%	36%
II. Organization	6	5%	30%
III. Form and Style	4	4%	16%
IV. Mechanics	5	4%	20%

To equalize the classifications it was necessary to go over 1% at sections I and III. This caused the total score to be 102%; however, the .02 of the total did not adversely affect the outcome of the grading.

The scale was designed to use the percentage of the classifications as the multiplier for the determination of the total for each of the classifications. By adding the 4 products together, a total, representing the final score, was computed. This figure then needed

RESEARCH PAPER EVALUATION STUDY																
OBJECTIVE GRADING SCALE																
GRADER NUMBER _____				PAPER NUMBER _____												

1. CONTENT				A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
A. Purpose and Direction				13	12	11	10	9	8	7	6	5	4	3	2	1
B. Choice of Subject				13	12	11	10	9	8	7	6	5	4	3	2	1
C. Use of Sources				13	12	11	10	9	8	7	6	5	4	3	2	1
D. Intelligent Discussion				13	12	11	10	9	8	7	6	5	4	3	2	1
E. Logic of Argument				13	12	11	10	9	8	7	6	5	4	3	2	1
F. Clarity of Interpretation				13	12	11	10	9	8	7	6	5	4	3	2	1
Total Points _____				X 6 = Total Content _____												
2. ORGANIZATION				A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
A. Use of Sources to Support Purposes				13	12	11	10	9	8	7	6	5	4	3	2	1
B. Writing to a Thesis				13	12	11	10	9	8	7	6	5	4	3	2	1
C. Effective Introduction				13	12	11	10	9	8	7	6	5	4	3	2	1
D. Effective Conclusion				13	12	11	10	9	8	7	6	5	4	3	2	1
E. Paragraph Development and Organ.				13	12	11	10	9	8	7	6	5	4	3	2	1
F. Consistent Point of View				13	12	11	10	9	8	7	6	5	4	3	2	1
Total Points _____				X 5 = Total Organization _____												
3. FORM AND STYLE				A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
A. Outline				13	12	11	10	9	8	7	6	5	4	3	2	1
B. Footnotes				13	12	11	10	9	8	7	6	5	4	3	2	1
C. Bibliography				13	12	11	10	9	8	7	6	5	4	3	2	1
D. Neatness and Appearance				13	12	11	10	9	8	7	6	5	4	3	2	1
Total Points _____				X 4 = Total Form and Style _____												
4. MECHANICS				A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
A. Diction				13	12	11	10	9	8	7	6	5	4	3	2	1
B. Punctuation				13	12	11	10	9	8	7	6	5	4	3	2	1
C. Spelling				13	12	11	10	9	8	7	6	5	4	3	2	1
D. Clear Gramm. Correct Sentences				13	12	11	10	9	8	7	6	5	4	3	2	1
E. Effective Sentence Structure				13	12	11	10	9	8	7	6	5	4	3	2	1
Total Points _____				X 4 = Total Mechanics _____												
-----				-----												
Total of 4 items _____				÷ 100 (move decimal 2 places to left) = Final Points _____												
-----				-----												
A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F	FINAL GRADE _____			
13	12	11	10	9	8	7	6	5	4	3	2	1	_____			

Figure 1. Research paper evaluation study - objective grading scale.

to be divided by 100 to yield the final numerical grade equivalent. A conversion table appeared at the end of the grade sheet to aid the rater in assigning a letter grade.

With the completion of the scale and the college professor validation, the first two stages of the research were completed. The use of the scale was then tested. The procedure for testing the scale involved 6 graders. These graders were all English majors with experience in subjective grading methods but very little, if any experience in the use of an objective grading instrument. On initial contact the graders were asked to help in a research project in which they would grade 50 library research papers. After gaining their consent the following materials were delivered to the graders:

- 1) Instructions to Graders (Appendix G)
- 2) Set #1 Research Papers (25 copies)
- 3) Set #2 Research Papers (25 copies)
- 4) Packet #1 Subjective Score Sheets (25 copies)
(Appendix H)
- 5) Packet #2 Objective Grading Scale Sheets (25 copies)
(Appendix I)
- 6) 2 red marking pens

The process followed was to grade 25 papers by a teacher subjective method and to assign a letter grade equivalent to the 1-13 number scale. After a period of 2 weeks, the raters graded the same 25 papers using the prepared objective scale. The papers and grade sheets were then collected for analysis. Because of the

bulk of the graded papers and grade sheets, they were not included in this paper but were retained by the researcher.

After the data were collected a correlation coefficient was computed to determine the degree of reliability of the two methods. The t-test for two related samples and the analysis of variance were two tests used to help determine the reliability of the objective scale. Computer analysis was used and the accompanying tables show the data that were obtained from that analysis.

CHAPTER IV

ANALYSIS OF THE DATA

After the raters were finished with the grading, the material was prepared for computer analysis. Two tests were used to test the reliability of the objective grading scale. They were 1) t-test for two related samples, and 2) analysis of variance. Also included in the analysis were several tables for reference as to the individual rater scoring using the two methods.

Table 1 shows the Descriptive Survey of the Totals-Number Grade. There were six raters and twenty-five papers. The table illustrates a grade for each paper, both subjective and objective. The conversion scale appears at the bottom of the table.

Table 2 shows the Descriptive Survey of the Totals-Letter Grade for the six raters on the twenty-five papers.

For a visual representation of the relationship between the two variables a scatter diagram (15, p. 75) was used. Figure 2 is a graph illustrating the relationship between the two variables for the entire group of 300 papers. The numbers on the horizontal and vertical axis represent the grade of A+ = 13 to F = 1. Figure 2 represents positive but less than perfect correlation.

The scatter diagrams for the six graders were included for a comparison, and they were all positively correlated with varying degrees of perfection (Figures 3, 4, 5, 6, 7, and 8).

Table 2

Descriptive Survey of the Totals - Letter Grade

	R-1	R-1	R-2	R-2	R-3	R-3	R-4	R-4	R-5	R-5	R-6	R-6
Paper No.	Sub.	Obj.	Sub.	Obj.	Sub.	Obj.	Sub.	Obj.	Sub.	Obj.	Sub.	Obj.
1	A-	A-	C	B-	B-	B+	A-	A-	B+	B+	B-	B+
2	D+	B+	C-	C	D+	B-	D-	D	B	C+	C	B
3	B-	B+	C	C+	C-	C	D-	D	B+	C	B+	B+
4	B-	B	B-	B-	B+	A-	B+	A-	C-	B-	C+	B-
5	C+	A-	B-	C+	B-	B-	B	B-	C+	B-	B+	A
6	B+	B+	B-	B	B+	A-	A	A-	C+	C	C+	B
7	A	B	C-	C	B-	B-	B-	B+	A-	C	B	A-
8	C-	C-	D	C-	D	C-	D-	D+	C+	D+	B+	A-
9	A	B	B+	B	A-	B-	C	B+	B+	B+	B-	B+
10	A-	B+	B	B	C-	B-	C	C+	A-	B+	B	A-
11	B	C+	B-	B-	C-	C	A-	A-	C+	B	B	A-
12	A+	A-	B+	B+	A	A-	A	A-	B-	B	B+	B+
13	B-	B-	C	C-	D+	C-	C-	C-	B-	A-	B	A-
14	A-	B-	C	C+	C+	C-	D	D+	B+	A-	C+	B
15	C	C	D-	C-	D	D	C	C+	D-	B+	B-	A-
16	B+	C+	B-	B-	C-	D	C+	C-	C-	B-	B-	B+
17	B-	B	D	C-	C-	C	D	C-	D+	B	C	B+
18	A	B+	D	C-	D+	D+	D-	C-	B	B	C+	B+
19	B	A-	C+	C	C-	C-	C-	C-	B+	A-	A-	A
20	C+	B+	D	D+	F	F	F	F	A-	B+	B	A-
21	B	C-	C	C-	C-	C+	C	C	D	C-	C+	B
22	A-	A-	A	B+	B-	B	B+	C+	B+	B+	B+	A-
23	B	B+	B-	C+	C+	B	C+	B	B	B	B	A-
24	B+	C+	C-	D+	C+	D	C	C-	B	B+	B	A
25	A	A-	B	B+	A-	B+	A-	B	A-	B+	B+	A-

**THIS BOOK
CONTAINS
NUMEROUS PAGES
WITH DIAGRAMS
THAT ARE CROOKED
COMPARED TO THE
REST OF THE
INFORMATION ON
THE PAGE.**

**THIS IS AS
RECEIVED FROM
CUSTOMER.**

SUBJECTIVE	13										1			
	12								2	2	4			
	11					1		2	1	5	4	1		
	10					1	3		1	7	8	1		
	9				1		2	1	3	3	7	1		
	8						2	6	5	6	2			
	7			1	1	2	2		2	6	2	1		
	6					3	2	4	1	1	2			
	5			1	1	4	5	1	3					
	4				1	1		1		1	1			
	3			1	2	6								
	2			1	2	2					1			
	1	2												
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
OBJECTIVE														

Figure 2. Positive correlation of the objective and subjective methods of grading research papers.

SUBJECTIVE	13										1		
	12								2	1	1		
	11							1		1	2		
	10						2			1			
	9				1		1			1	1		
	8							1	2	1			
	7									1	1		
	6						1						
	5				1								
	4									1			
	3												
	2												
	1												
	0	1	2	3	4	5	6	7	8	9	10	11	12
OBJECTIVE													

Figure 3. Grader one - positive correlation of the objective and subjective methods of grading research papers.

SUBJECTIVE	13													
	12									1				
	11													
	10								1	1				
	9								1	1				
	8						2	3	1					
	7					1								
	6				2		2	1						
	5			1		2								
	4													
	3			1	3									
	2				1									
	1													
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
OBJECTIVE														

Figure 4. Grader two - positive correlation of the objective and subjective methods of grading research papers.

SUBJECTIVE	13													
	12										1			
	11							1		1				
	10										2			
	9													
	8							2	1	1				
	7			1		1			1					
	6													
	5			1		1	3	1	1					
	4				1	1			1					
	3			1		1								
	2													
	1	1												
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
OBJECTIVE														

Figure 5. Grader three - positive correlation of the objective and subjective methods of grading research papers.

SUBJECTIVE	13													
	12										2			
	11								1		2			
	10						1				1			
	9							1						
	8									1				
	7				1				1					
	6				1	1	2			1				
	5				2									
	4													
	3				1	1								
	2			2	1	1								
	1	1												
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
OBJECTIVE														

Figure 6. Grader four - positive correlation of the objective and subjective methods of grading research papers.

SUBJECTIVE	13													
	12													
	11						1				3			
	10						1				3	2		
	9							1		2	1			
	8									1		1		
	7				1		1		1	1				
	6													
	5								2					
	4									1				
	3					1								
	2										1			
	1													
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
OBJECTIVE														

Figure 7. Grader five - positive correlation of the objective and subjective methods of grading research papers.

SUBJECTIVE	13													
	12													
	11											1		
	10									2	3	1		
	9										6	1		
	8									3	1			
	7							1	3	1				
	6								1	1				
	5													
	4													
	3													
	2													
	1													
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
OBJECTIVE														

Figure 8. Grader six - positive correlation of the objective and subjective methods of grading research papers.

The Grader Analysis (Table 3) illustrates the variance between the methods of grading for each of the raters. Variable No. represents the six raters and the four columns represent the two methods of grading with the appropriate Mean Scores and Standard Deviation Scores.

Table 3
Grader Analysis of Mean and Standard Deviation
Scores for Objective and Subjective Methods

Variable No.	Subjective Mean	Objective Mean	Subjective STDV	Objective STDV
1	9.280	9.000	2.289	1.855
2	6.560	6.920	2.515	1.874
3	6.440	6.800	2.787	2.757
4	6.520	6.960	3.442	2.905
5	8.120	8.640	2.551	1.916
6	8.520	10.400	1.360	1.020

The Paper Analysis (Table 4) illustrates the mean and standard deviation scores for the twenty-five papers. The t-test for two related samples (15, p. 170) was used to test the hypothesis. The test, with $df = 120$, yielded that the t-statistic is significant with a 2-tailed test at the .01 level of significance. The objective score is statistically higher than the subjective score, but the difference is regarded as having no practical significance.

Table 4

Papers Analysis of Mean and Standard Deviation
for Objective and Subjective Methods. (*) Sug-
gests disagreement among raters

Variable No.	Subjective Mean	Objective Mean	Subjective STDV	Objective STDV
1	9.000	10.000	1.826 *	1.000
2	5.000	7.167	2.160	2.267 *
3	6.833	7.167	2.853 *	2.192
4	8.000	9.167	1.732 *	1.344
5	8.167	9.000	1.067	1.826 *
6	9.000	9.333	1.826 *	1.700
7	8.833	8.333	2.267 *	1.886
8	5.000	5.667	2.769 *	2.427
9	9.500	9.333	1.979 *	0.745
10	8.500	9.167	2.291 *	1.344
11	8.167	8.667	1.863	1.886 *
12	10.833	10.333	1.675 *	0.745
13	6.667	7.500	1.795	2.693 *
14	7.333	7.333	2.625 *	2.357
15	4.500	7.000	2.291	2.769 *
16	7.167	6.833	1.772	2.267 *
17	4.833	7.333	1.772	2.055 *
18	6.167	7.167	3.532 *	2.544
19	7.833	8.333	2.339	3.037 *
20	5.333	6.167	3.902	4.298 *
21	6.000	6.167	1.826 *	1.463
22	10.167	9.667	1.213	1.374 *
23	8.167	9.167	0.898	1.213 *
24	7.667	6.833	1.795	3.236 *
25	10.667	10.167	0.943 *	0.687

Table 5
Summary Data for T-Test for Two Related Samples

	Subjective Scores	Objective Scores
Mean	7.573	8.120
Standard Deviation	2.810	2.541
Degree of Freedom = 149	T-Statistic = 3.350	

The second test was the Analysis of Variance (15, pp. 230-236). This test was an appropriate statistical procedure to use with two or more samples. It is mathematically equivalent to the t-test (15, p. 230).

Table 6
Summary Table for the Analysis of Variance - Objective

Source of Variation	df	Sum of Squares	Mean Squares	F-Ratio
Papers	24	269.8086	11.2420	3.1911
Graders	5	269.2773	53.8555	15.2869
Interaction	120	422.7578	3.5230	XXXXXX
Total	149	961.8438	XXXXXX	XXXXXX

The objective analysis is calculated as follows:

$$\frac{MS_{\text{papers}} - MS_{\text{interaction}}}{MS_{\text{papers}}} = \frac{11.2420 - 3.5230}{11.2420} =$$

$$\frac{7.7190}{11.2420} = .68662$$

Table 7

Summary Table for the Analysis of Variance - Subjective

Source of Variation	df	Sum of Squares	Mean Squares	F-Ratio
Papers	24	478.3281	19.9303	4.6880
Graders	5	188.2070	37.6414	8.8540
Interaction	120	510.1602	4.2513	XXXXXX
Total	149	1176.6950	XXXXXX	XXXXXX

The subjective analysis is calculated as follows:

$$\frac{MS_{\text{papers}} - MS_{\text{interaction}}}{MS_{\text{papers}}} = \frac{19.9303 - 4.2513}{19.9303} =$$

$$\frac{15.6790}{19.9303} = .78669$$

The F-statistic for the subjective is higher than for the objective, thus the null hypothesis is retained that no significant difference existed between the two methods of grading research papers.

CHAPTER V

SUMMARY AND CONCLUSIONS

As previously stated the purpose of this study was to develop an evaluative tool to aid the teacher of composition in grading research papers. This tool was an objective scale designed to be weighted to produce more objective grades on the research papers. The following considerations were in the foreground of the present research:

- 1) What correlations would there be between a paper graded by the traditional subjective method and an objective scale?
- 2) On an average which method would yield the higher scores?
- 3) Which method was more reliable?
- 4) Was the original hypothesis retained or rejected?

One interesting conclusion that came out of this research involved the correlation between the two methods of grading. It was revealed in reviewing the literature that more reliable instruments were needed by teachers of composition for the purpose of grading composition papers. It was assumed for this study that this instrument, most appropriate for this purpose, would be an objective scale. And by using this scale one could expect better reliability of grading. This was not necessarily true. Collectively, the rater's scores on the two methods (Figure 2, p. 20) correlated positively. The correlation was, however, not perfect. It was previously assumed that there

would be very little correlation, but the scatter graph illustrated the opposite. With all but six or seven scores, the correlation was a 3 point scale correlation. It was definitely established that in this study the two methods correlated. In viewing Figures 3-8 (pp. 21-25), the individual graders scores were shown to correlate. They tended to vary in regard to the degree of correlation, but all showed positive correlation. Some were more nearly perfect than the others.

It was found that the scores on the papers that were graded with the objective scale were higher than the scores of the papers graded by the subjective method. The mean scores (Table 3, p. 27) for graders 2-6 were higher using the objective scale. Grader 1 had a higher mean on the scores of the subjective method. The comparison of subjective and objective mean scores indicated fairly definite correlation between the two methods.

In this research the hypothesis was concerned with the establishment of reliability for one of the two methods of grading. Two tests were used for this purpose. The results of the two tests did not yield any practical significant difference between the two methods of grading the research papers. On the t-test for two related samples, the \bar{t} was statistically significant, but for practical purposes, the difference was not significant. The analysis of variance test used indicated that the subjective statistic was higher, but it did not establish reliability on either the subjective method or the objective scale.

The null-hypothesis was retained--what difference, if any, will there be in the reliability of grades on student term papers if graded by a subjective method or by an objective scale?

In summary:

- 1) the two methods do correlate positively.
- 2) graders tend to score papers higher using the objective scale.
- 3) significant reliability could not be established for the objective scale or for the subjective method of grading.

In regard to the present research, even though reliability was not established, certain implications were drawn. It seems that the need for an objective method of grading research papers still exists. With the amount of time involved in evaluating a paper of this length, some criteria for grading papers of this type would be advantageous to assure the student of the most objective assessment possible.

One reason for the high level of correlation between the two methods of grading might have been the degree to which some graders use the subjective method of grading. With the shorter themes and essays the graders might not think out the items for evaluating as carefully as they would on a longer research paper. This higher level of sophistication in determining the grades on papers without the use of a scale might be the indicating factor for the high degree of correlation between the two methods. The scores obtained in using

the subjective method seemed to be very closely correlated with the scores obtained in grading the papers by the objective scale.

Another problem in scale research, confronting the researcher, was the interpretation of the scale items by the selected raters. This was evident in looking over the comment sheets. The graders viewed some of the items differently, and they placed more emphasis on certain items even though the scale had already been weighted.

Because of the results of the present study, certain implications for future research were seen. It has been established in the review of the literature that a scale for objective evaluation of papers is needed. Future research needs to be conducted in the development of a reliable scale and its testing. One method for conducting this research might be the design of having one paper with fifty graders. Each grader would be asked to grade the paper using the subjective method and, after a lapse of time, grade the same paper using an objective scale. This would yield a correlation of grades on the same paper. This correlation would also be available within the papers using the same method. If the correlation was higher on the objective instrument one would have a more reliable instrument for the evaluation of research papers.

In conclusion, objective methods of evaluation need to be developed for the teachers of English composition. Even though correlation exists between the two methods, there exists a need for a more objective and reliable method.

BIBLIOGRAPHY

BIBLIOGRAPHY

1. Braddock, Richard, Lloyd-Jones, Richard and Schoer, Lowell.
Research in Written Composition. Champaign, Illinois:
National Council of Teachers of English, 1963.
2. David, Brother Clermat.
"Rating High School Themes," Catholic Education Review, 60:
174-8, March, 1962.
3. Diedrich, P. B.
"Making and Using Tests," English Journal, 44:135-40, 1955.
4. Denby, R. V.
"NCTE-ERIC Report - Composition Evaluation," English Journal
57:1215-21, November, 1968.
5. Follman, John C. and Anderson, James A.
"An Investigation of the Reliability of Five Procedures for
Grading English Themes," Research in the Teaching of English,
pp. 190-200, Fall, 1967.
6. Fostvedt, D. R.
"Development of a Scale for Evaluation of High School Student
Themes," Montana Education, 39:20-1, November, 1962.
7. Godshalk, Fred I., Swineford, Frances and Coffman, William E.
The Measurement of Writing Ability. New York: College
Entrance Examination Board, 1966.
8. Hudelson, Earl.
English Composition, Its Aims, Methods, and Measurements.
Bloomington, Illinois: Public School Publishing Company, 1923.
9. Huddleston, Edith M.
"Measurement of Writing Abilities at the College Level:
Objective vs. Subjective Testing Techniques," Journal of
Experimental Education, 22:165-213, March, 1954.
10. Jewett, Arno and Bish, Charles E.
Improving English Composition. Chicago: NEW-Dean Langmuir
Project on Improving English Composition, 1965.
11. Judine, Sister M.
A Guide for Evaluating Student Composition. Champaign,
Illinois: National Council of Teachers of English, 1965.

12. Lewallen, Lawrence Jay.
Bern High School, Bern, Kansas 66408.
13. Morrill, A. Reed.
"Evaluating Term Papers," Improving College and University Teaching, 14:181-2, Summer, 1966.
14. Newbery, R. A.
"Objective Indices in the Assessment of Essays," British Journal of Educational Psychology, 37:403-5, November, 1967.
15. Roscoe, John T.
Fundamental Research Statistics for the Behavioral Sciences.
New York: Holt, Rinehart and Winston, Inc., 1969.
16. Schumann, P. F.
"What Criteria Do We Use in Grading Compositions?" English Journal, 57:1163-5, November, 1968.
17. Turner, A. L.
"Theories and Practices in Evaluation and Correction of Oral and Written Compositions," Peabody Journal of Education, 19:266-273, 1942.

APPENDIXES

APPENDIX A

NAMES OF PROFESSORS

Dr. Marjorie Adams Associate Professor of English	Dr. Paul McCarthy Associate Professor of English
Dr. Joye Ansdell Associate Professor of English	Dr. Richard McGhee Assistant Professor of English
Dr. William Brondell Associate Professor of English	Dr. Will Moses Professor of English
Dr. Walter Eitner Associate Professor of English	Dr. Benjamin Nyberg Assistant Professor of English
Mr. William Evans Assistant Professor of English	Dr. Charles Pennel Associate Professor of English
Mr. Vincent Gillespie Assistant Professor of English	Dr. Helen Petrullo Assistant Professor of English
Mrs. Esther Glenn Assistant Professor of English	Dr. John Rees Assistant Professor of English
Dr. David Houser Assistant Professor of English	Mr. Harold Schneider Assistant Professor of English
Dr. William Hummel Professor of English	Dr. Mary Schneider Assistant Professor of English
Dr. Kenneth Johnston Assistant Professor of English	Mr. Donald Stewart Assistant Professor of English
Mr. William Koch Assistant Professor of English	Dr. Mary Frances White Associate Professor of English
Mr. Russell Laman Assistant Professor of English	

APPENDIX B

FIRST PROFESSOR RESPONSE LETTER

Box 82
Bern, Kansas
May 26, 1970

Dr. Marjorie Adams
Associate Professor of English
Kansas State University
Manhattan, Kansas

Dear Dr. Adams:

I am presently completing my Master's degree in the College of Education at Kansas State University. For my thesis I have chosen to work on the evaluation of research papers written by high school seniors. It is my intent to develop a scale for a more objective examination and evaluation of these papers rather than the usual subjective teacher evaluation.

To establish validity for the scale that will be developed, I am seeking the assistance of university English professors at Kansas State University in terms of their ideas as to the items important in the grading of research papers. It would be appreciated if you would take the time to complete the enclosed sheet and return it in the stamped self-addressed envelope.

Thank you for your time and guidance in this work.

Sincerely,

Jay Lewallen

APPENDIX C

PROFESSOR RESPONSE SHEET

If you were grading a research paper that included the following items:

- 1) 2000-2500 words in length
- 2) library sources (primary and secondary)
- 3) footnotes
- 4) bibliography
- 5) no assigned subject

what would be your criteria for grading? Please arrange the items in order of importance and assign percentage values for weighting the items.

Thank you.

APPENDIX D

SECOND PROFESSOR RESPONSE LETTER

Box 82
Bern, Kansas
July 8, 1970

Dr. Marjorie Adams
Associate Professor of English
Kansas State University
Manhattan, Kansas

Dear Dr. Adams:

In a continuing process toward the development of a workable evaluation scale, the following scale is being submitted to you for further refinement and percentage weighing of the various elements. From responses of Kansas State University English professors to my May 26th letter, this scale was developed. Accompanying the scale is a sheet that gives the criteria for the development of such a scale. The purposes and assignment are more clearly defined there than they were on the first letter.

After examining all of the material would you please assign percentage weights to the various parts of the scale. If some of the items on the scale need, in your opinion, to be changed, feel free to add or subtract any item. Your assistance in this work is greatly appreciated.

Please send your remarks in the self-addressed stamped envelope at your earliest convenience. Thank you.

Sincerely,

Jay Lewallen

APPENDIX E

CRITERIA FOR DEVELOPMENT OF SCALE

PURPOSES OF THE PAPER:

1. to further develop writing in a communicative manner.
2. to extend mechanical correctness of English to the longer composition.
3. to develop the paper from a selected thesis.
4. to introduce the proper methods of gathering secondary source materials.
5. to synthesize the material in a logical planned manner.
6. to introduce the use of quotations and footnote form.
7. to introduce the correct compilation of bibliographical form.

ASSIGNMENT OF THE PAPER:

1. 2000 to 2500 words in length
2. library sources
3. method of taking notes
4. footnotes
5. bibliography
6. no assigned subject

APPENDIX F

RESEARCH PAPER EVALUATION SCALE-PERCENTAGE RESPONSE

1. CONTENT % _____
 - 1.1 Purpose and Direction
 - 1.2 Choice of Subject
 - 1.3 Use of Sources
 - 1.4 Intelligent Discussion
 - 1.5 Logic of Argument
 - 1.6 Specifics to Support Generalizations
 - 1.7 Clarity of Interpretation

2. ORGANIZATION % _____
 - 2.1 Use of Sources to Support Purposes
 - 2.2 Writing to a Thesis
 - 2.3 Effective Introduction
 - 2.4 Effective Conclusion
 - 2.5 Paragraph Development and Organization
 - 2.6 Consistent Point of View

3. FORM AND STYLE % _____
 - 3.1 Title Page
 - 3.2 Outline
 - 3.3 Footnotes
 - 3.4 Bibliography
 - 3.5 Neatness and Appearance

4. MECHANICS % _____
 - 4.1 Diction
 - 4.2 Punctuation
 - 4.3 Spelling
 - 4.4 Clear Grammatically Correct Sentences
 - 4.5 Effective Sentence Structure

APPENDIX G

RESEARCH PAPER EVALUATION STUDY

INSTRUCTIONS TO GRADERS

1. Enclosed you will find two sets of papers. Set #1 will be graded first and set #2 will be graded second.
2. The packet of score sheets are enclosed in envelopes and should be opened one set at a time.
3. Grade the first set of papers using the subjective method and list your comments and circle the grade on each score sheet. (Refer to the sample copy of score sheet.)
4. Be sure to place your number and paper number at the top of each score sheet.
5. Feel free to make markings on the research papers with the pencils that are provided.
6. After all papers in set #1 are graded, place all score sheets in the packet and set aside with the research papers in set #1.
7. A lapse of time of approximately 2 weeks should be observed before grading the second set of papers.
8. The same procedure should be followed in grading set #2. The papers are the same research papers.
9. In using the objective scale, there are four sections. Circle the items in each section then figure the section score and complete the scale form. (Refer to the sample copy of score sheet.)
10. For reference the grade equivalent is placed at the top of the column of numbers.
11. When you are finished with both sets, contact me and I will pick up the two sets of research papers and the two sets of grade sheets.

APPENDIX H

RESEARCH PAPER EVALUATION STUDY

SUBJECTIVE GRADING SHEET

GRADER NUMBER _____ PAPER NUMBER _____

COMMENTS:

A+	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F	FINAL GRADE _____
13	12	11	10	9	8	7	6	5	4	3	2	1	

AN OBJECTIVE SCALE FOR
EVALUATING ENGLISH RESEARCH PAPERS

by

LAWRENCE JAY LEWALLEN

A. B., Washburn University, 1964

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1971

The teaching and evaluating of composition has received considerable criticism in the past few years. Much of this criticism is justified because the field is almost void of any uniform method of grading that helps to eliminate the subjectiveness of grading student work. A scale for the grading of library research papers is needed.

The purpose of this research was to determine what difference, if any, will there be in the reliability of grades on student term papers if graded by a subjective method or by an objective scale. These basic considerations were part of the study: 1) selection of criteria for a scale, 2) weighing of the criteria, 3) development of an objective scale, 4) establishment of scale validity, and 5) analysis of reliability.

The previous research on this topic has been primarily concerned with the evaluation of short essays and compositions. The scaling of items for evaluation of research papers has been sparse.

The development of the scale began with responses from college English professors to determine items important for evaluating research papers. From these responses an objective scale was developed and weighted. The research papers for this study were written by 25 first semester junior college freshmen. To test the reliability of the scale, six high school English teachers were asked to grade one set of 25 papers using some subjective method and, after a lapse of time, to grade the same 25 papers using the objective scale.

After the data were collected a correlation coefficient was computed to determine the degree of reliability of the two methods. Two tests were used: 1) t-test for two related samples and 2) analysis of variance. The t-test, with $df = 120$, yielded that the t-statistic was significant with a two-tailed test at the .01 level of significance. The objective score was statistically higher, but the difference was regarded as having no practical significance. The F statistic for the subjective was higher than for the objective, on the analysis of variance test, thus the null hypothesis was retained that no significant difference existed between the two methods of grading research papers.

The following conclusions were drawn from the research:

- 1) the two methods do correlate positively.
- 2) graders tend to score papers higher using the objective scale.
- 3) significant reliability could not be established for the objective scale nor for the subjective method of grading.

The need for objective methods of evaluating student papers still exists. Even though correlation exists between the two methods, future research needs to be conducted to develop the scale that will produce more reliable and objective evaluation of student research papers.