EFFECTS OF TWO TYPES OF TEACHER RESPONSE TO ESSAYS UPON TWELFTH GRADE STUDENTS' GROWTH IN WRITING PERFORMANCE

Dissertation for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
MARIE E. KELLEY
1973

This is to certify that the

thesis entitled

EFFECTS OF TWO TYPES OF TEACHER RESPONSE TO ESSAYS UPON TWELFTH GRADE STUDENTS' GROWTH IN WRITING PERFORMANCE

presented by

Marie E. Kelley

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Secondary Education and Curriculum

Elizabeth H. Rusk
Major professor

Date July 20, 1973

O-7639

El

T

or not th

response

in the ar

of twelft

response

permitte

in growt

formance

one acad

between

drafts a

careful

is a ma

Experts

vided v

ABSTRACT

EFFECTS OF TWO TYPES OF TEACHER RESPONSE TO ESSAYS UPON TWELFTH GRADE STUDENTS' GROWTH IN WRITING PERFORMANCE

By

Marie E. Kelley

The purpose of the study was to determine whether or not the use of two distinct types of teacher written response to student essays produced significant differences in the amount of growth in writing performance of two groups of twelfth grade students who received differing teacher response treatments. Data related to three hypotheses permitted the reporting of findings comparing differences in growth in general, mechanical, and total writing performance for two groups of students (1) over the period of one academic semester, (2) during the period of time between contiguous units of study, and (3) between initial drafts and revisions of essays.

Teachers of written composition have assumed that careful evaluation of student essays combined with revision is a major factor in the development of writing skills. Experts in the teaching of written composition have provided various kinds of advice related to the ways teachers

ought to e
often been
Experiment
amounts an
contradict
combined c
the effect
determine
students'
knowledge
further i
and isola
students'

S
twelfth g
which the
school.
one group
essays t
directiv
rated by
ending r
formanc
written
experim
revisio
Student

ought to evaluate written composition. This advice has often been based upon intuition rather than sound research. Experimental investigations of the effects of different amounts and types of teacher evaluation have resulted in contradictory findings. Many of these experiments have combined other variables in the composition process with the effects of teacher evaluation, and it is difficult to determine the relative impact of specific variables upon students' growth in writing performance. The state of knowledge about composition evaluation may be advanced with further investigation of the effects of clearly specified and isolated kinds of teacher responses to essays upon students' growth in writing performance.

Subjects for the experiment were twenty-eight twelfth grade students in a one-semester composition class which the author taught in a metropolitan Nebraska high school. Students were divided randomly into two groups; one group of students received clarifying responses for essays they wrote, and the other group of students received directive responses. Three pretest and post-test essays rated by a team of independent raters provided entering and ending measures of the quality of students' writing performance. Students received one of the two types of written teacher response for eight essays during the experimental period, and ratings for these essays and their revisions were included in the data for the experiment. Student t-tests were computed for forty-two comparisons of

interest

confiden

signific

conclusi

1.

2.

3.

interest for the three hypotheses. The .05 level of confidence was established as the criterion required for significance.

From the findings of this experiment the following conclusions were reached:

1. There is no assurance that a teacher's choice of either clarifying responses or directive responses to student essays produces an advantage which influences the amount of growth in general, mechanical, or total writing performance of twelfth grade students during the course of an academic semester.
2. There is no conclusive evidence that the use of either the clarifying or directive teacher response produces an advantage which consistently influences the amount of growth in writing performance which occurs between contiguous units of study, although the clarifying response may be more effective than the directive response when used in an exposition unit immediately followed by an argumentation unit.
3. There is no conclusive evidence that a teacher's choice of either the clarifying or directive response on initial drafts of essays produces an advantage which consistently influences the amount of growth in writing performance of twelfth grade students on revisions, but the clarifying response

may be more effective than the directive response for producing growth on revisions of expository essays.

4. There is no assurance, when a twelfth grade student revises an essay after the teacher has used the clarifying or directive response, that the revised essay will show evidence of better writing performance than that shown on an initial draft of an essay, although net growth scores for revisions over a total semester may show a slight advantage for the use of the clarifying response.

EFFECTS

ESS

in

Depar

EFFECTS OF TWO TYPES OF TEACHER RESPONSE TO ESSAYS UPON TWELFTH GRADE STUDENTS' GROWTH IN WRITING PERFORMANCE

By

Marie E. Kelley

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Secondary Education and Curriculum

1973

The aut
major professor
of her advisory
and Dr. Louise
of this dissert
Special
Schools for gra
to the author'
School who coo
which made thi
The au
Dr. Allen Ditt
long hours rat
of this study
The a
husband and s

ACKNOWLEDGMENTS

The author wishes to express gratitude to her major professor, Dr. Elizabeth Rusk, and the other members of her advisory committee, Dr. Dale Alam, Dr. George Price, and Dr. Louise Sause, for their guidance in the preparation of this dissertation.

Special thanks also go to the Lincoln Public Schools for granting permission to conduct this study and to the author's colleagues at Lincoln Northeast High School who cooperated in the arrangement of the conditions which made this study possible.

The author is indebted to Dr. Larry Andrews, Dr. Allen Dittmer, and Mrs. Mary Commers, who spent many long hours rating the essays which formed the data base of this study.

The author also expresses appreciation to her husband and son for their encouragement and assistance.

LIST OF TABLES

Chapter

I. INTRODU

State
Signi
Hypot
Limit
Assur
Defin
Desig
Overv

II. RELATED

Relia
Facto
Wri
Evalu
and
Alter
on
Concl

III. PROCED

Sour
Proc
The
The
Anal
Trea

TABLE OF CONTENTS

Table

LIST OF TABLES

Table

The te
subject to ass
publication o
Writing[1] some
cerned than t
of learning p
write well in
the United St
of various me
frequency of
writing pract
composition.
Applebee[2] poi
effort in the

[1] Rep
(Ann Arbor:
1970).

[2] Jam
English Inst
Crofts, 1968

CHAPTER I

INTRODUCTION

The teaching of written composition has been subject to assessment for many years, and since the 1970 publication of the results of the National Assessment of Writing[1] some school districts have been even more concerned than they were in the past with providing the kinds of learning programs which will prepare their students to write well in comparison with other students and adults in the United States. Investigators have studied the effects of various methods of instruction in composition, of frequency of writing, and of alternatives to actual writing practice upon student performance in written composition. Investigations such as that of Squire and Applebee[2] point out the immense expenditure of time and effort in the teaching of writing in the schools. Yet,

---

[1] Report 3, 1969-70 Writing: National Results (Ann Arbor: National Assessment of Educational Progress, 1970).

[2] James Squire and Roger K. Applebee, High School English Instruction Today (New York: Appleton-Century-Crofts, 1968).

researcher

express di

students.

English
general
much mo
on the
effecti
theory

A g

the teachin

student essa

teaching of

various type

performance

constitutes

time. Write

have suggest

advised that

each paper,

student's wr

the teacher

student's wr

have encoura

on papers.

---

[3] Her

Holt, Rineha

researchers, teachers, and the general public continue to express dissatisfaction with the writing performance of students. Muller observed:

> English teachers go on as if they agreed with the general public that composition is absolutely essential, much more important and practical than literature; yet, on the record nothing in English is taught less effectively, amid more confusion and conflict of theory or hunch.[3]

A great proportion of the time teachers devote to the teaching of written composition is spent in evaluating student essays. Many recent investigations into the teaching of composition have centered on the impact of various types of evaluation of composition upon the writing performance of students in an effort to determine what constitutes effective use of the teacher's evaluation time. Writers who have attemped to answer this question have suggested widely varying approaches. Some have advised that the teacher engage in intensive evaluation of each paper, making copious notes on all aspects of the student's written performance; others have suggested that the teacher select and comment on one element of the student's writing performance in each essay; still others have encouraged teachers to write only comments of praise on papers.

---

[3]Herbert J. Muller, The Uses of English (New York: Holt, Rinehart, and Winston, 1967), p. 98.

Recent
compositions ha
ments accordin
the effects of
on student wri
research is ne

The sp
was whether o
used by the t
combined with
cantly differ
mance of two

Many
students' wr
essays after
assumption i
identificati
are most lik
This study f
written teac
to determine
other in the
writing skil

Recent analyses of teachers' comments on students' compositions have resulted in the categorization of comments according to types, but few investigations exploring the effects of controlled use of selected types of comments on student writing performance have been initiated. More research is needed in this area of composition evaluation.

Statement of the Problem

The specific question investigated in this study was whether or not two distinct types of written response used by the teacher in the evaluation of student writing, combined with student revision of essays, produced significantly differing amounts of growth in the writing performance of two groups of twelfth grade students.

Significance of the Study

Many experts have assumed that the improvement of students' writing skills is influenced by the revision of essays after the teacher has evaluated them. If this assumption is correct, teachers would be aided by the identification of the kinds of evaluative responses which are most likely to lead to improvement in writing skills. This study focused on the use of two specific modes of written teacher response to student essays in an attempt to determine whether or not one mode is preferable to the other in the facilitation of improvement in students' writing skills.

study.

I.

II.

III.

Investigatio

regarding th

improvement

reported inc

Hypotheses of the Study

Three major hypotheses were investigated in this study. Stated in the null form, these hypotheses are:

I. When pretest and post-test scores of total, general, and mechanical writing performance, as measured by the Diederich composition rating scale, are compared by the use of a student t-test, there is no significant difference between the growth in writing performance of the group of students who received clarifying responses for their written compositions and the growth in writing performance of the group of students who received directive responses for their written compositions.

II. When scores of total, general, and mechanical writing performance on initial compositions for one unit of study, as measured by the Diederich composition rating scale, are compared by use of a student t-test with scores of total, general, and mechanical writing performance on initial compositions for a contiguous unit of study, there is no significant difference between the growth in writing performance of the group of students who received clarifying responses for their written compositions and the growth in writing performance of the group of students who received directive responses for their written compositions.

III. When scores of total, general, and mechanical writing performance on initial drafts of written compositions, as measured by the Diederich composition rating scale, are compared by use of a student t-test with scores of total, general, and mechanical writing performance on revisions of these compositions, there is no significant difference between the amount of change in these scores for the group of students who received clarifying responses for their written compositions and the group of students who received directive responses for their written compositions.

Investigation of the above hypotheses resulted in findings regarding the relationship of the two modes of response to improvement in writing performance. The relationships reported include growth in writing performance over a

period of
assignment
rent, and
written re

Th
were limit
composed e
further li
written re
time. Var
of that re
compositio
some of th
extending
able to on
Restrictio
was conduc
impossible

Th
evaluation
accompanie
factor in

period of time, growth between the initial draft of one assignment and the initial draft of the succeeding assignment, and the effects of two different modes of teacher written response upon student revision efforts.

Limitations of the Study

The study and the applicability of its findings were limited by the use of only one composition class composed entirely of high school seniors. The study was further limited by the investigation of only two types of written responses. A third limitation was one imposed by time. Varying lengths of time for research of the nature of that reported in this study are suggested by experts in composition research, and this study's duration fell within some of the recommended parameters; however, a study extending for a full academic year would have been preferable to one which extended for only a single semester. Restrictions imposed by the setting in which the research was conducted made the extension of the time period impossible.

Assumptions of the Study

The major assumption of this study was that careful evaluation of student compositions by the teacher, accompanied with student revision or rewriting, is a major factor in the development of writing skills.

Two
gated in th
written com
student re
which he i
clarifying
student ev
alternativ
expected t
response a
major aspe
of ideas,

Th
this study
Three term
tant to f
study. T
is the wr
the score
scale. G
sition is
for the f
rating sc
Diederich
cated a h

Definition of Terms

Two modes of written teacher response were investigated in this study. The directive response is a brief written comment which gives a specific direction to the student regarding the improvement of the writing skills which he is expected to demonstrate in his writing. The clarifying response is a question designed to help the student evaluate the nature of his ideas and consider alternatives in relation to the writing skills he is expected to demonstrate in his writing. Both the directive response and the clarifying response were focused upon major aspects of the composition process, e.g., development of ideas, organization, wording, and flavor.

The Diederich composition rating scale was used in this study to measure the writing performance of students. Three terms related to this composition scale were important to facilitating the reporting of the findings of this study. Total writing performance for a written composition is the writing performance which is measured by the sum of the scores for the eight factors of the Diederich rating scale. General writing performance for a written composition is that which is measured by the sum of the scores for the first four factors of the Diederich composition rating scale (ideas, organization, wording, flavor). Diederich, after use of factor analysis procedures, indicated a high correlation among these four factors, and

described

writing pe

a written

of the sco

scale (usa

Diederich

mechanical

copy of th

component

Sources of

Or

Lincoln Nc

the fall s

as the sou

conduct th

Schools.

enrolled i

writing as

tuted the

---

[4] P

Evaluation

ed. by Fre

for Superv

p. 198.

described the combined factors as a measure of general writing performance.[4] Mechanical writing performance for a written composition is that which is measured by the sum of the scores for the last four factors of the Diederich scale (usage, punctuation, spelling, handwriting). Diederich identified these combined factors as a measure of mechanical writing performance. Appendix A contains a copy of the Diederich scale and definitions of its eight component factors.

Design of the Study

Sources of the Data

One composition class which the author taught at Lincoln Northeast High School, Lincoln, Nebraska, during the fall semester of the 1972-1973 academic year was used as the source of the data for this study. Permission to conduct the study was granted by the Lincoln Public Schools. Of the thirty-two twelfth grade students initially enrolled in the class, the twenty-eight who completed all writing assignments during the experimental period constituted the experimental sample.

---

[4]Paul B. Diederich and Frances R. Link, "Cooperative Evaluation in English," Evaluation as Feedback and Guide, ed. by Fred T. Wilhelms (Washington, D.C.: Association for Supervision and Curriculum Development, NEA, 1967), p. 198.

Collection of

Prior

a value-clari

formulate cla

response migh

to students'

test the eff

formance. T

weeks. The

experimenta

selection p

directive r

experimenta

responses

period. A

students w

modes of d

first week

for maximu

response,

essays and

period du

Th

the experi

competenci

At the end

he had dev

Collection of Data

Prior to the experiment, the investigator attended a value-clarification workshop in which she learned to formulate clarifying responses. She believed the clarifying response might be used effectively in teachers' responses to students' essays; thus she designed an experiment to test the effects of its use upon students' writing performance. The experimental period extended for seventeen weeks. The investigator assigned the members of the experimental class to two groups through the use of random selection procedures. One group of students received directive responses to all essays written during the experimental period; the other group received clarifying responses to all essays written during the experimental period. As a measure of their entering writing performance, students wrote three essays, representing three different modes of discourse, on three different days during the first week of the experimental period. In order to provide for maximum student learning from the two modes of teacher response, the teacher wrote comments on these pretest essays and had students revise each of them in a class period during the second week of the experimental period.

The teacher taught five composition units during the experimental period. The units were designed to teach competencies in writing in different modes of discourse. At the end of each unit, each student wrote an essay which he had developed during the prewriting instructional

period. Th
wrote eithe
depending u
assigned at
After teach
essays in c

Dur
students wr
ments simil

All
during the
were writte
essays and

Analysis a
the Data

At
gator trai
in the stu
cation and
English an
versity of
of the ess
chairman o
in Lincoln
essay exam
the rating
five or mo

period. The teacher evaluated the students' essays and wrote either directive or clarifying responses on them, depending upon the groups to which the students had been assigned at the beginning of the experimental period. After teacher evaluation, the students revised their essays in class.

During the final week of the investigation period students wrote three essays in response to three assignments similar to those used for pretest assignments.

All essays and revisions which the students wrote during the experimental period were photocopied after they were written. The investigator then coded the photocopied essays and stored them.

Analysis and Treatment of the Data

At the end of the experimental period, the investigator trained three raters to rate the 532 essays included in the study. Two of the raters hold doctorates in education and professorial rank in both the Department of English and Department of Secondary Education at the University of Nebraska-Lincoln. These two raters rated all of the essays included in the study. The third rater, chairman of the English department of another high school in Lincoln, has had experience in rating College Board essay examinations. She rated only those essays for which the ratings of the first two raters were discrepant by five or more points. Diederich recommended a procedure

similar to th
After two tra
ent rating o
per week. Th
structions t
they should
manner as to
guard agains
which the st
arranged the
for each of
any effects
rater fatig

Aft
the investi
ratings us
coefficien
test essay
also prepa
data neede
gation.
the Unive
three nul
.05 was e

similar to this in Evaluation as Feedback and Guide.[5] After two training sessions, the raters began the independent rating of the essays at the rate of two sets of essays per week. The investigator had provided uniform instructions to the raters relative to the rating procedures they should follow and had arranged the papers in such a manner as to provide for the anonymity of students and to guard against the raters knowing the point in time during which the students had written any of the essays. She also arranged the papers in each set in a different random order for each of the two major raters, in order to guard against any effects in the ratings which might be attributed to rater fatigue.

After all essays included in the study were rated, the investigator computed reliability coefficients for all ratings using the Pearson product moment correlation coefficient, and mean scores for both the pretest and post-test essays of all students included in the study. She also prepared summary sheets containing the additional data needed for testing the hypotheses in this investigation. Then, with the use of the computer facilities at the University of Nebraska, the student t-tests for the three null hypotheses were computed. An alpha level of .05 was established as the level of statistical

---

[5] Ibid., pp. 200-208.

significan

of this st

Th

In

described,

hypotheses

overview o

Li

compositio

reviewed i

In

procedures

experiment

and treatm

procedures

Al

shown in o

also repor

In

marized, a

reviewed.

presented.

significance required for the rejection of the hypotheses of this study.

Overview of the Organization of the Study

This study is divided into five chapters.

In Chapter I, the basic problem has been defined, described, and limited by the formulation of three specific hypotheses. Important terms have been defined, and an overview of the design of the study has been presented.

Literature related to the evaluation of written composition and the use of teacher responding behaviors is reviewed in Chapter II.

In Chapter III, a description of the instructional procedures used by the author during the course of the experiment is provided. Procedures used for the collection and treatment of the data are described, and training procedures used to prepare the team of raters are presented.

All steps used in the analysis of the data are shown in Chapter IV. The major findings of the study are also reported in Chapter IV.

In Chapter V, the findings of the study are summarized, and their limitations and applicability are reviewed. Recommendations for further research are presented.

The
composition
concern. Or
measures.
be achieved
effects of
writing per
is determin
which best
in the eval
sideration
reinforceme
in evaluati
students?
provide a f
evaluation

CHAPTER II

RELATED LITERATURE

The literature about teacher evaluation of written composition seems to cluster into three major areas of concern. One of these is reliability of composition rating measures. If reliable ratings of student writing cannot be achieved, then the teacher has no way of determining the effects of his evaluative procedures upon students' writing performance. Another important area of interest is determination of the aspects of the composition process which best lead to student growth when they are emphasized in the evaluation process. A third major area is consideration of teachers' evaluative responses as types of reinforcement. Do different types of reinforcement used in evaluative responses produce varying effects in students? In this chapter, these three areas of concern provide a focus for the review of literature related to evaluation of composition.

The de

the quality of

throughout the

of ratings sho

has noted, ann

usually report

century, a num

composition e

Willing[4] deve

Monroe[5] and G

---

[1]Rich

Schoer, Resea

National Cou

[2]E.

Writing by Y

XI (June, 19

[3]M.

Ability in

College Rec

[4]M.

sition," En

[5]W

ments of A

Yearbook o

cation, Pa

Company,

6

Compositi

ed. by W.

383-95.

Reliability of Composition Rating Measures

The development of reliable methods for measuring the quality of essays has been a question of concern throughout the century. The importance of the reliability of ratings should not be underestimated, since, as Braddock has noted, annual gains in writing performance are small, usually reported as about 5 per cent.[1] Early in the century, a number of persons attempted to construct composition evaluation scales. Thorndike,[2] Hillegas,[3] and Willing[4] developed some of the better known scales. Monroe[5] and Greene[6] criticized the unreliability of such

---

[1]Richard Braddock, Richard Lloyd-Jones, and Lowell Schoer, Research in Written Composition (Champaign, Ill.: National Council of Teachers of English, 1963), p. 7.

[2]E. L. Thorndike, "A Scale of Merit in English Writing by Young People," Journal of Educational Psychology, XI (June, 1911), 361-68.

[3]M. B. Hillegas, "A Scale for the Measurement of Ability in English Composition by Young People," Teachers College Record, XII (September, 1912), 331-84.

[4]M. H. Willing, "Measurement of Written Composition," English Journal, VII (March, 1918), 193-202.

[5]W. S. Monroe, "The Unreliability of the Measurements of Ability in Written Composition," in Twenty-second Yearbook of the National Society for the Study of Education, Part I (Bloomington: Public School Publishing Company, 1923), pp. 169-71.

[6]Harry A. Greene, "English Language, Grammar, and Composition," in Encyclopedia of Educational Research, ed. by W. S. Monroe (New York: Macmillan, 1950), pp. 383-95.

scales, and P

use of such s

Meck

develop comp

of writing, s

that these s

One such han

is A Guide f

Some

to establish

writing have

In a study s

Diederich, P

in student w

grading. T

at three di

---

[7]H.
ment and Ev
pp. 233-36.

[8]He
and Literat
by N. L. Ga
p. 987.

[9]Si
Evaluating
Council of

[10]Pa
Carlton, Fa
Bulletin 61
1961).

scales, and Remmers and Gage[7] concluded that widespread use of such scales by classroom teachers is not feasible.

Meckel identified several recent attempts to develop composition scales and handbooks for the analysis of writing, such as the California Essay Scale, but noted that these scales have not been statistically tested.[8] One such handbook which includes several composition scales is A Guide for Evaluating Student Composition.[9]

Some of the recent researchers who have attempted to establish a uniform procedure for the rating of student writing have approached the problem more scientifically. In a study sponsored by Educational Testing Service, Diederich, French and Carlton[10] investigated the qualities in student writing which cause readers to differ in their grading. Three hundred papers written by college freshmen at three different institutions were read by readers from

---

[7]H. H. Remmers and N. L. Gage, Educational Measurement and Evaluation (New York: Harper and Brothers, 1943), pp. 233-36.

[8]Henry C. Meckel, "Research on Teaching Composition and Literature," in Handbook of Research on Teaching, ed. by N. L. Gage (Chicago: Rand McNally & Company, 1963), p. 987.

[9]Sister M. Judine, I. H. M., ed., A Guide for Evaluating Student Composition (Champaign, Ill.: National Council of Teachers of English, 1965).

[10]Paul B. Diederich, John W. French, and Sydell T. Carlton, Factors in Judgments of Writing Ability, Research Bulletin 61-15 (Princeton: Educational Testing Service, 1961).

six different
instructions
the papers; t
their general
disliked in t
fewer than fi
The correlati

An ir
in the study
revealed eig
readers: ide
punctuation,
developed a
on a one to
being highes
double weigh
base for thi
emphasize th
scale includ
for each of
the Diederic
included in

11P
Improving E
Charles Bis
Association

12P
Writing Abi

six different academic disciplines. These readers had no instructions concerning what to look for in evaluating the papers; they arranged them into nine levels representing their general merit and commented on things they liked and disliked in the papers. None of the 300 papers received fewer than five grades, and 101 received all nine grades. The correlation among all readers was .31.[11]

An investigation of written comments of the readers in the study conducted by Diederich, French, and Carlton revealed eight primary factors which influenced the readers: ideas, organization, wording, flavor, usage, punctuation, spelling, and handwriting. Diederich developed a scale in which each of these factors is ranked on a one to five scale with one being lowest and five being highest. The idea and organization factors are given double weight. Diederich stated that there is no research base for this practice, but that "teachers usually want to emphasize the importance of these two qualities."[12] The scale includes definitions of high, middle, and low ratings for each of the eight factors. A complete reproduction of the Diederich scale, including these definitions, is included in Appendix A of this study.

---

[11]Paul B. Diederich, "Grading and Measuring," in Improving English Composition, ed. by Arno Jewett and Charles Bish (Washington, D.C.: National Education Association, 1965), pp. 83-84.

[12]Paul B. Diederich, "How to Measure Growth in Writing Ability," English Journal, LV (April, 1966), 446.

As Di

his scale, he

are possible

evaluative c

Evaluation a

compositions

third reader

ratings are

scale, usual

The

methods of r

a rating pr

mended the

rather than

unreliabili

by standard

examiners."

Smi

determine w

student wri

13 P

Operative E

and Guide,

Association

NEA, 1967),

14 B

Methods of

Journal of

As Diederich experimented with the application of his scale, he established the fact that reliable ratings are possible when raters are trained to use identical evaluative criteria. The rating procedure he outlined in Evaluation as Feedback and Guide, in which four student compositions are rated by two independent raters, with a third reader rating papers for which the first two readers' ratings are variant by more than ten points on the Diederich scale, usually yields high reliability coefficients.[13]

The conclusions of Cast in a 1939 study of various methods of rating compositions support the development of a rating procedure such as that of Diederich. Cast recommended the analytic method of evaluating compositions rather than the general impression method and stated that unreliability of rating "can evidently be greatly reduced by standardized instructions and by the training of examiners."[14]

Smith developed and administered a test designed to determine whether or not judgment in the evaluation of student writing can be measured validly, efficiently, and

---

[13]Paul B. Diederich and Frances R. Link, "Co-operative Evaluation in English," Evaluation as Feedback and Guide, ed. by Fred T. Wilhelms (Washington, D.C.: Association for Supervision and Curriculum Development, NEA, 1967), p. 201.

[14]B. M. D. Cast, "The Efficiency of Different Methods of Marking English Composition, Part II," British Journal of Educational Psychology, X (February, 1940), 59.

reliably. He
consisted of f
students. Exp
these samples
was then admi
nonteachers.
judgment betw
however, that
the experts i
test.[15]

The r
they are eval
Cast. There
scribing how
it may be as
the writing
untrained as

In a
rater diffe
constant wh
found no si
raters. He

reliably. He prepared two forms of the test, each of which consisted of five short samples of writing from fifth grade students. Experts in the teaching of composition ranked these samples on the basis of overall quality. The test was then administered to various groups of teachers and nonteachers. Smith found no significant differences in judgment between nonteachers and teachers. He did find, however, that more than half of the teachers differed with the experts in judgment of the writing samples in the test.[15]

The results obtained by Smith are not surprising if they are evaluated in terms of the recommendations made by Cast. There were no instructions to the subjects describing how they should rate the writing samples; thus, it may be assumed that no consistent analytic analysis of the writing samples existed. Also, the subjects were untrained as raters.

In an experiment designed to determine whether rater differences measured by the Smith test would remain constant when raters rated another set of essays, Whalen found no significant differences between "good" and "bad" raters. He concluded: "The results cast doubt on the

---

[15]Vernon H. Smith, "Measuring Teacher Judgment in the Evaluation of Written Composition," Research in the Teaching of English, III (Fall, 1969), 191.

validity of S

assessing ess

Dona

to the studi

to develop a

sition. He

submitted th

requesting t

importance.

(2) Develop

(5) Organiz

graphing.

agreement a

order to es

themes, ran

grade stude

the validat

ratings fai

teria. Fo

English co

16

for Measuri

Paper prese

cational Re

ERIC No. ED

"Composition

Journal, LX

17

of High Sch

cational Re

validity of Smith's test as a general instrument for assessing essay-rating behavior."[16]

Donald Fostoedt conducted an investigation similar to the studies of Diederich. The purpose of his study was to develop a set of criteria for evaluating written composition. He selected five criteria from the literature and submitted them to college and high school English experts, requesting that the criteria be ranked according to their importance. These criteria were: (1) Coherence and Logic, (2) Development of Ideas, (3) Diction, (4) Emphasis, and (5) Organization Through Sentence Structure and Paragraphing. "Validity of the criteria was established by agreement at the 1 per cent level in the ranking."[17] In order to establish reliability of the criteria, twenty themes, randomly selected from 256 themes written by twelfth grade students, were graded by thirty teacher experts using the validated criteria. A statistical analysis of these ratings failed to establish the reliability of the criteria. Fostoedt concluded that "although teachers of English composition may feel that criteria are important

---

[16]Thomas E. Whalen, "A Validation of the Smith Test for Measuring Teacher Judgment of Written Composition," Paper presented at the Annual Meeting of the American Educational Research Association (New York, February, 1971), ERIC No. ED 049 275, summarized in Danial J. Dieterich, "Composition Evaluation: Options and Advice," English Journal, LXI (November, 1972), 1268.

[17]Donald R. Fostoedt, "Criteria for the Evaluation of High School English Composition," The Journal of Educational Research, LIX (November, 1965), 110.

in evaluati

in the empl

gest, howev

established

school were

rating them

Post

side the re

the evaluat

use a scal

situation

rating sca

ratings.

V

specific

Briggs c

handwrit

marking

in many

it as a

raters

total e

influen

work.

in evaluating themes, there is no evidence of consistency in the employment of such criteria."[18] Fostoedt did suggest, however, that reliability for the criteria could be established if English experts from the same region or high school were oriented to the use of the criteria before rating themes.

Fostoedt's study points to problems which lie outside the realm of the development of a reliable scale for the evaluation of compositions. Raters must be trained to use a scale consistently. Also, aspects of the rating situation which are unrelated to the factors within the rating scale itself may influence the reliability of the ratings.

Various studies have investigated the effects of specific factors upon the assessment of writing. Dennis Briggs conducted such a study in which he concluded that handwriting style significantly influenced teachers' marking of essays.[19] Although handwriting is not included in many composition rating scales, Diederich did include it as a factor in his composition scale. Perhaps, if raters recognize handwriting as a small component of the total essay, they may be less inclined to allow it to influence their judgment of other aspects of the written work.

---

[18] Ibid., p. 111.

[19] Dennis Briggs, "The Influence of Handwriting on Assessment," Educational Research, XIII (November, 1970), 55.

Sever

Florida condu

rating time a

themes. Two

reproduced wi

members of or

essay; the me

analysis of

"no signific

face condit

to test dir

both groups

of test th

researcher

readings;

grade bas

the secor

They als

leniency

to them

the ra

outcor

face

in t

Several investigators at the University of South Florida conducted a study concerned with the effects of rating time and differing typeface upon the grading of themes. Two groups of ten raters graded eight themes reproduced with a variety of four styles of typeface. The members of one group were allowed one minute to rate each essay; the members of the second, two minutes. Statistical analysis of the data collected in this experiment revealed "no significant differences associated with time and typeface conditions."[20] Although the study was not designed to test directly the leniency effect, it was noted that both groups of raters awarded higher grades in the set of test themes rated last. On the basis of this study, the researchers recommended that raters give each theme two readings; the first, rapid reading for the assignment of a grade based upon an overall impression of the theme and the second, reading for correction of mechanical errors. They also cautioned that raters should be sensitive to the leniency effect and guard against awarding higher grades to themes graded later.[21]

Leniency in the ratings of papers rated later in the rating period is only one of a number of possible outcomes of rater fatigue. Braddock suggested that

---

[20]J. C. Follman, et al., "Effects of Time and Typeface on Level and Reliability of Theme Grades," Research in the Teaching of English, IV (Spring, 1970), 54.

[21]Ibid., pp. 56-57.

fatigue may also cause raters to become more severe or erratic in their ratings. In order to guard against this, he suggested the avoidance of lengthy rating sessions and recommended regular rest periods for the raters.[22]

Braddock identified a further rater variable--the possibility that the personal feelings of a rater could influence his ratings. To guard against this in experimental studies, Braddock stressed the importance of the anonymity of the writer. The rater should not know the identity of the writer, the treatment the writer received in the experiment, or the time when the paper was written.[23]

Another factor which research has shown to influence the reliability of essay grades is the number of essays written by a single student. Studies conduced by Coffman[24] and the College Entrance Examination Board[25] have demonstrated that the reliability of the assessment of student performance increases when more than one essay written by the same student is rated. Kincaid investigated a problem somewhat related to the findings of these two

---

[22]Braddock, Lloyd-Jones, and Schoer, op. cit., p. 11.

[23]Ibid., p. 10.

[24]William E. Coffman, "On the Validity of Essay Tests of Achievement," Journal of Educational Measurement, III (1966), 156.

[25]Fred I. Godshalk, et al., The Measurement of Writing Ability (New York: College Entrance Examination Board, 1966), p. 39.

studies: doe

response to a

a valid basis

concluded:

> An evalu
> teaching
> ability"
> from eac
> of disco
> procedur
> post-tes
> similar

Some

obtaining r

through the

objective t

findings o

the advisa

study cond

the combi

writing a

dicted by

in the u

---

ations
Ed.D. d
State U
Lloyd-

of the
Psycho

studies: does a single paper written by a student in response to a given topic at a particular time constitute a valid basis for evaluation of writing performance? He concluded:

> An evaluation of "the effectiveness of different teaching methods with students who vary in writing ability" should be based on several samples of writing from each student, writing on different topics (modes of discourse) and during different days, the same procedure being followed both with the pretest and the post-test themes, the nature of the assignment being similar if not the same on both these occasions.[26]

Some researchers have bypassed the problem of obtaining reliable measures of student writing performance through the evaluation of written essays by using various objective tests to measure writing performance. The findings of Vernon and Millican, however, cast doubt on the advisability of this practice. They concluded from a study conducted at the London Institute of Education that the combined judgment of two or more raters "does yield a writing ability factor which can be only partially predicted by tests."[27] Braddock, too, pointed out limitations in the use of objective measures of writing performance in

---

[26]Gerald L. Kincaid, "Some Factors Affecting Variations in the Quality of Students' Writing" (Unpublished Ed.D. dissertation, [Michigan State College] Michigan State University, 1953), summarized in Richard Braddock, Lloyd-Jones and Schoer, op. cit., p. 95.

[27]P. E. Vernon and G. D. Millican, "A Further Study of the Reliability of English Essays," British Journal of Psychology, VII (November, 1954), 73.

his statemer
attempt to r
Afte
development
formance, or
reliable me
the use of
related to
are control
several stu
reliability

Th
factors i
indicates
these st
sition t
to which
sized a
a compo
must do
meanin

his statement that objective tests "make little or no attempt to measure the larger elements of composition."[28]

After studying the research concerned with the development of reliable measures of student writing performance, one might conclude that the achievement of reliable measurement is possible if raters are trained in the use of a common set of criteria and if variables related to the raters, the writers, and the topics assigned are controlled. Within this context, Braddock cited several studies which obtained highly significant reader reliability coefficients.[29]

Factors of Emphasis in Evaluation of Written Composition

The previously cited literature concerned with the factors involved in the judgment of writing performance indicates how writing is judged by various raters. While these studies provide valuable information for the composition teacher, they do not provide definite guidelines as to which aspects of writing performance should be emphasized as he writes comments on the essays of students in a composition class. Most experts agree that the teacher must do more than judge an essay if evaluation is to be a meaningful part of the total process of learning in a

---

[28]Braddock, Lloyd-Jones and Schoer, op. cit., p. 45.

[29]Ibid., p. 14.

composition c

concerned wit

upon motivati

writing. A p

skills are mo

instructiona

the process

which aspect

when he eva

utilize the

for the stu

this quest

Ir

students'

use of a

mechanica

structio

the crit

caution

individ

Theme
reprin
ed. by
Natio

Harco

composition class. A survey of recent articles and books concerned with the teaching of writing reveals an emphasis upon motivation and instruction preceding the act of writing. A pertinent question is which of the writing skills are most effectively taught in the prewriting instructional period, and which skills are taught through the process of teacher evaluation? More specifically, which aspects of composition should the teacher emphasize when he evaluates essays if he is to most effectively utilize the evaluation process as a learning experience for the student? A survey of the literature concerned with this question reveals contradictory opinions.

In 1955, Dusel suggested limiting comments on students' essays to responses to ideas. He advised the use of a separate tally card for recording each student's mechanical errors and the assignment of remedial instruction based on these errors.[30] Lou La Brant advised the criticism of ideas, form and structure, with a note of caution that criticism should be given with regard to the individual student's ability to use it.[31] The Commission

---

[30]William J. Dusel, "Some Semantic Implications of Theme Correction," English Journal, XLIV (1955), 390-97, reprinted in A Guide for Evaluating Student Composition, ed. by Sister M. Judine, I. H. M. (Champaign, Ill.: National Council of Teachers of English, 1965), p. 45.

[31]Lou L. LaBrant, We Teach English (New York: Harcourt, Brace, 1951), p. 183.

on English,

marking of

Ideally
and mar
comment
tive an
might b
success

Recognizing

provide thi

students, t

attending t

time.[33] p

intensive

High s
every
good
is th
writi

toward t

marginal

errors.

most he

Englis
1965),

Compo

on English, in its 1965 report, recommended intensive marking of essays:

> Ideally the teacher should not only read every paper and mark its formal errors but should write detailed comment. The comment . . . should always be constructive and specific, showing the student exactly what might be done to improve the theme and what has been successfully done in the theme as presented.[32]

Recognizing that the teacher does not have enough time to provide this "ideal" kind of evaluation for all essays of students, the Commission suggested the alternative of attending to one or two aspects of composition at a time.[33] Paul Diederich, however, stated opposition to intensive evaluation:

> High School English teachers who painstakingly mark every mistake on a theme are doing more harm than good . . . the most common effect of such correcting is that "most students learn to hate and fear writing."[34]

Many writers agree that a terminal comment directed toward the larger elements of composition should supplement marginal symbols used to indicate mechanical writing errors. Albert Kitzhaber advised that these comments are most helpful if they are "specific, pointed, and

---

[32]Commission on English, Freedom and Discipline in English (New York: College Entrance Examination Board, 1965), p. 99.

[33]Ibid., p. 101.

[34]"Corrections Hurt Writers, Expert Says," College Composition and Communication, XIV (May, 1963), 84.

constructive
comment shou
essay, since
structural c
Squi
hensive stud
United State
they observe
composition
identified
ideas, clea
important.
17 per cent
teach writ
whole cycl
Wh
when they
the questi
Most
sponsc

3
The Teach
Book Comp

3
on Studer
220.

English
Crofts.

constructive."[35] James Green suggested that the terminal comment should be based upon the thesis statement of the essay, since it "makes the most important semantic and structural commitments. . . ."[36]

Squire and Applebee, as a part of their comprehensive study of high school English instruction in the United States, asked department chairmen in the schools they observed which criteria for the evaluation of written composition were most important. The department chairmen identified organization and appropriate development of ideas, clear thinking, and sentence structure as most important. Yet, Squire and Applebee noted that only in 17 per cent of the schools were comments "designed to teach writing and thinking--the avowed purpose of the whole cycle of writing, correcting, and revising."[37]

What elements of composition do teachers stress when they evaluate student essays? Janet Emig answered the question this way:

> Most of the criteria by which students' school-sponsored writing is evaluated concerns the accidents

---

[35]Albert Kitzhaber, Themes, Theories, and Therapy: The Teaching of Writing in College (New York: McGraw-Hill Book Company, 1963), p. 135.

[36]James L. Green, "A Method for Writing Comments on Student Themes," English Journal, LVII (February, 1968), 220.

[37]James Squire and Roger K. Applebee, High School English Instruction Today (New York: Appleton-Century-Crofts, 1968), p. 121.

rather t
spelling
than the
sophist

Another inve

teacher res

analysis of

number and

essays, Mar

New York hi

student ess

ations, and

ten and twe

they differ

She also fo

teachers' c

the identi

to the lar

obtained r

Th

seem to in

Graders:
National

Analysis
of Averag
Ed.D. dis
Buffalo,
71-28,040

> rather than the essences of discourse--that is, spelling, punctuation, penmanship, and length rather than thematic development, rhetorical and syntactic sophistication, and fulfillment of intent.[38]

Another investigator sought to determine the emphasis of teacher responses to student writing through scientific analysis of actual teacher comments. In a survey of the number and kinds of teacher markings placed on student essays, Marshall analyzed 168 essays collected from three New York high schools. She concluded that markings on student essays consist primarily of symbols and abbreviations, and that comments on papers of students in grades ten and twelve differ in nature (although the ways in which they differed were not consistent across the three schools). She also found a positive result for her hypothesis that teachers' comments on student essays consist primarily of the identification of technical errors, rather than response to the larger elements of composition, although the obtained result was not statistically significant.[39]

The findings of Squire and Applebee and Marshall seem to indicate that the effectiveness of comments

---

[38]Janet Emig, The Composing Processes of Twelfth Graders: NCTE Research Report No. 13 (Urbana, Ill.: National Council of Teachers of English, 1971), p. 93.

[39]Barbara Ruth Cohn Marshall, "A Survey and Analysis of Teachers' Markings on Selected Compositions of Average Students in Grades 10 and 12" (Unpublished Ed.D. dissertation, State University of New York at Buffalo, 1971; Ann Arbor: University Microfilms, No. 71-28,040), p. 48.

written on e

investigati

year colleg

Confrey cla

their effec

concluded t

students in

One

written on

teacher mus

Dusel studi

time necess

teach writi

essay whic

would spen

Squire and

their stud

papers.[42]

4

Made by Fo
Education

Henry C. H
Literature
Gage (Chi

4

Teaching
XLIII, No

4

written on essays has not changed much since the 1927 investigation of Confrey. In a study involving forty first year college English instructors in eleven colleges, Confrey classified comments written on essays according to their effectiveness in promoting improved writing and concluded that less than 20 per cent of the comments aided students in improving their writing.[40]

One might assume that the inadequacy of comments written on essays is related to the amount of time a teacher must spend in order to write helpful comments. Dusel studied this time factor and concluded the evaluation time necessary if a teacher is to give suggestions which teach writing and thinking is 8.6 minutes for each 250 word essay which is evaluated. Thus, a teacher with 150 students would spend about 21.5 hours evaluating a set of papers.[41] Squire and Applebee observed that the average teacher in their study spent nine to twelve hours per week correcting papers.[42]

---

[40]A. Confrey, "An Investigation of the Comments Made by Forty Instructors Upon Students' Themes," Catholic Education Digest, XXIV (1927), 335-40, summarized in Henry C. Meckel, "Research on Teaching Composition and Literature," Handbook of Research on Teaching, ed. by N. L. Gage (Chicago: Rand McNally and Co., 1963), p. 986.

[41]William J. Dusel, "Determining an Efficient Teaching Load in English," Illinois English Bulletin, XLIII, No. 1 (1955), 13.

[42]Squire and Applebee, op. cit., p. 69.

In s

teacher's t

Robert Denb

the area of

an "apparen

practices d

years."[43]

Re

of teache

for teach

evaluati

neglecte

emphasi

reinfor

these s

in the

and re

effec

of co

eval

mary

wer

of

Ev

In spite of the fact that a great proportion of the teacher's time is devoted to the evaluation of composition, Robert Denby stated that almost no progress has occurred in the area of composition evaluation, and that there has been an "apparent absence of change in teachers' evaluation practices despite research findings of the last ten years."[43]

Research studies concerned with evaluative responses of teachers have often failed to provide clear guidelines for teachers who wish to improve the quality of their evaluative responses to student writing. They have often neglected a definition of the aspects of composition emphasized in the evaluative process and the types of reinforcement used in the evaluative response. Also, these studies have sometimes been tied to other variables in the composition program, such as frequency of writing and revision; thus, it is difficult to determine specific effects which might be attributed to a particular method of composition evaluation.

In one of the early studies of composition evaluation, John Fellows tested the effects of the detailed marking of all errors. Two groups of ninth grade students were involved in the twelve-week study. For one group of students, teachers carefully evaluated all essays,

---

[43]Robert V. Denby, "NCTE/ERIC Report--Composition Evaluation," English Journal, LVII (November, 1968), 1215.

marked all mechanical errors, assigned a grade, and returned the papers to students for revision. Teachers read the compositions of students in the second group and assigned a grade, but indicated no errors and required no rewriting. An error count of pretest and post-test essays revealed no significant difference between the gains of students in the two groups, except perhaps for some of the brightest pupils.[44] This study was concerned with improvement in the mechanical aspects of writing, and corrections were made with the use of a code. One might assume, therefore, that no actual comments were written on student essays.

Several studies, such as those of Lokke and Wykoff,[45] Mark Christiansen,[46] and Dressel, Schmid, and Kincaid,[47] were concerned with gains of students in writing performance when the amount of theme writing is increased.

---

[44]John Ernest Fellows, "The Influence of Theme-Reading and Theme-Correction on Eliminating Technical Errors in the Written Compositions of Ninth Grade Pupils," University of Iowa Studies in Education, VII, No. 1 (1932), p. 43.

[45]Virginia Lokke and George Wykoff, "Double Writing in Freshman Composition--An Experiment," School and Society, XLVIII (November, 1948), 437-39.

[46]Mark Christiansen, "Tripling Writing and Omitting Readings in Freshman English: An Experiment," College Composition and Communication, XVI (May, 1965), 122-24.

[47]Paul Dressel, John Schmid, Jr., and Gerald Kincaid, "The Effect of Writing Frequency Upon Essay-Type Writing Proficiency at the College Level," Journal of Educational Research, XLVI (December, 1952), 285-93.

The findings
writing perf
number of wr
Maize[48] at
ences in pe
who wrote t
After analy
Allen obser
may have b
frequency
the effec

T
effects
Buxton c
with fre
train e
assigne
"Writi
except

Engl
Publ

of
Con
Pes
ve

The findings of these investigations indicated no gains in writing performance resulting from an increase in the number of writing assignments. A study conducted by Maize[48] at Purdue University did show significant differences in performance for an experimental group of students who wrote twice as many essays as a control group wrote. After analysis of the Maize study, however, Sutton and Allen observed that the differences in writing performance may have been due to instructional factors other than frequency of writing.[49] These studies gave no attention to the effects of teacher evaluation.

The Buxton study, however, was concerned with the effects of evaluation as well as writing frequency. Buxton conducted his study at the University of Alberta with freshman students enrolled in an emergency program to train elementary teachers. The students were randomly assigned to three groups. Papers of students in the "Writing" group were evaluated with no marks or grades except for "a paragraph of comment emphasizing the good

---

[48]Ray C. Maize, "A Study of Two Methods of Teaching English Composition to Retarded College Freshmen" (Unpublished Ph.D. dissertation, Purdue University, 1952).

[49]Joseph T. Sutton and Eliot D. Allen, "The Effect of Practice and Evaluation on Improvement in Written Composition," United States Office of Education Cooperative Research Project No. 1993 (Deland, Florida: Stetson University, 1964), p. 7.

qualities o
required to
"Revision"
errors mark
elements of
written on
to students
group did
but did no
groups.
groups wr
500 words
Both wri
pretest
indepen
"the st
revise
than e
pract

obse

Eff
St
dis
Un

qualities of the paper."[50] Students in this group were not required to revise their papers. Papers of students in the "Revision" group were evaluated thoroughly with mechanical errors marked, and with comments concerned with larger elements of composition and suggestions for improvement written on the essays. They were then graded and returned to students for revision. The students in the "Control" group did the regular assignments required for the courses, but did none of the writing required of the other two groups. Students in both the "Writing" and "Revision" groups wrote a total of sixteen papers of approximately 500 words each during the seven-month experimental period. Both written essays and objective measures were used as pretest and post-test measures of writing performance. Two independent evaluators rated the essays. Buxton found that "the students whose writing was thoroughly criticized and revised showed improvement that was significantly greater than either the control group that did no writing or the practice group that was not asked to revise."[51]

In his review of the Buxton study, Braddock observed that since the raters knew which essays were

---

[50]Earl W. Buxton, "An Experiment to Test the Effects of Writing Frequency and Guided Practice Upon Students' Skill in Written Expression" (Unpublished Ph.D. dissertation, Stanford University, 1958; Ann Arbor: University Microfilms, No. 58-3596), p. 27.

[51]Ibid., p. 81.

pretests a
ratings in
three grou
writing pe
also point
thorough e
Buxton's s
conducted
A
writing
revision
in a Mas
through
wrote a
student
classe
Period
Heys

pp

A

pretests and which were post-tests, the results of the ratings indicate the relative composition quality of the three groups but are not a precise measure of gains in writing performance from pretest to post-test. Braddock also pointed out that the relative influence of grades, thorough evaluation, and revision cannot be determined in Buxton's study and recommended that similar experiments be conducted with subjects of differing age levels.[52]

Another study which tested the effects of frequent writing combined with rigorous teacher correction and revision is that of Heys. The experiment involved students in a Massachusetts high school enrolled in grades nine through twelve. Every week students in the writing classes wrote a theme which the teacher corrected rigorously; then students revised these themes. Students in the reading classes wrote a theme every third week, and spent one period of each week reading books which they had chosen. Heys reported these findings:

> . . . except for some seniors (but not all) and except for some low groups (but not all) and except for the area of content and organization (but not always), we got consistently better results from those students in the reading classes.[53]

---

[52]Braddock, Lloyd-Jones, and Schoer, op. cit., pp. 65-70.

[53]Frank Heys, Jr., "The Theme-a-Week Assumption: A Report of an Experiment," English Journal, LI (May, 1962), 322.

As is true

Heys did nc

correction.

MeM

comments di

He defined

written the

or lack of

students r

assigned

mented up

assigned

(4) teac

fourth c

findings

proved

note th

levels

favour

abilit

their

---

of F

Them

Colu

As is true of many of the studies reviewed in this chapter, Heys did not clearly define what constituted "rigorous correction."

MeMechan investigated the effects of evaluative comments directed toward the "salient features" of essays. He defined these "salient features" as "those aspects of a written theme which stand out, either for their excellence or lack of it."[54] Four matched groups of grade eight students received one of four treatments: (1) teachers assigned letter grades to all themes, (2) teachers commented upon the salient features of all themes, (3) teachers assigned letter grades to one-fourth of the themes, and (4) teachers commented upon the salient features of one-fourth of the themes.[55] McMechan concluded, from the findings of his study, that none of the four methods proved more successful than the others. He did, however, note that when the students were sub-grouped into ability levels that " . . . high ability students progressed favourably regardless of marking procedures and that lower ability students appeared to progress favourably only if their work received regular teacher attention."[56]

---

[54]Melville Y. McMechan, "The Relative Effectiveness of Four Procedures for Evaluating Students' Written Themes" (Unpublished M.A. Thesis, University of British Columbia, 1961), p. 16.

[55]Ibid., p. 54.

[56]Ibid., p. 56.

Bur

in grades t

resulting f

In an artic

the finding

to be no mo

frequent w

that no ad

evaluatio

P

involved

twelve.

which su

produce

finding

in orde

give a

1/2 da

discus

Freque
Schoo
Unite
Proje

Are T
1964)

Effec

Burton and Arnold[57] conducted studies with students in grades ten to twelve to test improvement in writing resulting from frequent writing and intensive evaluation. In an article summarizing the studies, Lois Arnold reported the findings and concluded that intensive evaluation seemed to be no more effective than moderate evaluation; that frequent writing, in itself, does not improve writing; and that no advantage for frequent writing and intensive evaluation were found for any ability level.[58]

An investigation conducted by McColly and Remstad involved approximately 300 students in grades eight through twelve. Their findings contribute to the bulk of evidence which suggests that additional writing alone does not produce improvement in writing performance. Based on the findings of their study, the investigators recommended that in order to teach composition effectively, "teachers should give a weekly writing task on which they base about 2 and 1/2 days of practical explanation, student practice, discussion, revising, rewriting, etc."[59]

---

[57]Dwight L. Burton and Lois V. Arnold, "Effects of Frequency of Writing and Intensity of Evaluation Upon High School Students' Performance in Written Composition," United States Office of Education Cooperative Research Project 1523 (Tallahassee: Florida State University, 1963).

[58]Lois V. Arnold, "Writer's Cramp and Eyestrain--Are They Paying Off?" English Journal, LIII (January, 1964), 14.

[59]William McColly and Robert Remstad, "Comparative Effectiveness of Composition Skills Learning Activities in

Wil

evaluation i

writing peri

composition

identified

written com

strengths a

class peri

assignmen

niques in

class th

the othe

methods

used.

two te

superi

for ir

quest

agai

writ

the

Coo

of

ni

(C

in

William Clark experimented with techniques of evaluation in an attempt to assess their effects upon the writing performance of students enrolled in freshman composition at the United States Air Force Academy. He identified two basic tehniques: (1) the use of extensive written comments on an essay to inform the student of his strengths and weaknesses, and (2) discussion, during one class period, of two or three essays written for the assignment. Each of three instructors used these techniques in four different ways in his four classes. In one class the teacher used one of the two methods; in a second, the other; in a third class a combination of the two methods was used; and in the fourth class neither method was used. "No reliable evidence was found to indicate that the two techniques, used singly or in combination, were superior to instruction which offered students no guidance for improving their writing."[60]

The evidence from the studies which involve the question of frequency of writing seems overwhelmingly against the practice of merely increasing the number of writing activities. The findings of the studies concerned

---

the Secondary School," United States Office of Education Cooperative Research Project 1528 (Madison: University of Wisconsin, 1963), p. 65.

[60]William G. Clark, "An Evaluation of Two Techniques of Teaching Freshman Composition: Final Report" (Colorado Springs: Air Force Academy, 1968), summarized in Daniel J. Dieterich, op. cit., 1269.

with intens

The Buxton

intensive t

identified

treatment.

be given to

their effec

students.

In

Annotation

should be

effects o

improveme

sition a

which co

properti

ness an

ateness

"to pr

instru

resea

The s

Anno
Res
Yor

with intensity of evaluation, however, are contradictory. The Buxton study indicated positive effects resulting from intensive teacher evaluation, but all other studies identified no particular benefits resulting from this treatment. One might conclude that more attention should be given to the nature of teachers' evaluative comments and their effects, if any, upon the writing performance of students.

In a study entitled "The Dimensions of Composition Annotation," William McColly developed an instrument which should be helpful in developing research centered on the effects of teachers' evaluative responses as related to the improvement of student writing. He discovered that composition annotation possesses distinct internal properties which correspond to those of compositions. These general properties are: (1) Competence and Adequacy, (2) Helpfulness and Positiveness, (3) Appearance, and (4) Appropriateness.[61] One of the objectives of McColly's study was "to produce, or lead to the production of, a conceptual instrument whose use will make possible certain kinds of research on the dynamics of annotation as feedback."[62] The scale developed in this study seems to have merit for

---

[61]William McColly, "The Dimensions of Composition Annotation," United States Office of Education Cooperative Research Project S-216 (Oswego: State University of New York at Oswego, 1965), pp. 29-32.

[62]Ibid., pp. 33-34.

analyzing t

question o

be more th

Se

the effect

these resp

reinforce

effects o

attitude

reinfor

would a

and rew

failur

negati

oral a

mecha

infor

behav

Learn
p. 56

Stude
of So

analyzing the nature of teachers' comments so that the question of the effects of these comments on students can be more thoroughly researched.

Evaluative Responses as Reinforcement and Feedback

Several recent researchers have been concerned with the effects of teacher responses upon student behavior when these responses are categorized into different kinds of reinforcers. They have often sought to determine the effects of specific types of reinforcement upon student attitudes and achievement.

Hilgard and Bower studied the ideas concerned with reinforcement and concluded that most learning theorists would agree that positive reinforcements refer to successes and rewards, and negative reinforcements to punishments and failures.[63] Teachers typically provide both positive and negative reinforcement in the classroom as they provide oral and written feedback in response to student behaviors.

Solomon and Rosenberg defined feedback as "the mechanism by which a machine, device, or organism receives information about the nature and/or effects of its behavior."[64] They defined two variables within feedback--

[63]E. R. Hilgard and G. H. Bower, Theories of Learning (New York: Appleton-Century-Crofts, 1966), p. 563.

[64]Daniel Solomon and Larry Rosenberg, "Teacher-Student Feedback and Classroom Social Structure," Journal of Social Psychology, LXII (April, 1964), 197-98.

information

effects of

upon the so

that "if th

correct and

praise for

is correct,

students w

will be er

So

feedback;

to writte

Diederic

each ess

not more

improve

directe

gestion

errors

advanc

he can

expres

Journa

information and reinforcement--and speculated about the effects of kinds of oral feedback from teacher to student upon the social structure of the classroom. They stated that "if the teacher does not respond differently to a correct and an incorrect student contribution, but gives praise for the attempt to both, with no indication of which is correct, the differentiation between good and poor students will not be apparent to the class, and each group will be encouraged to participate."[65]

Solomon and Rosenberg spoke of the effects of oral feedback; yet, some experts would apply the same principles to written feedback in response to student essays. Paul Diederich advised that teachers find at least one thing in each essay that the student has done well and preferably not more than one thing the student should attempt to improve in his next paper. He recommended that comments directed toward aspects students should improve be suggestions rather than prescriptions. If teachers mark all errors, he said, "the student learns nothing; he only advances one step further toward a settled conviction that he can't write and there is no use trying."[66] Moffett expressed agreement with this philosophy, stating that

---

[65] Ibid., p. 203.

[66] Paul B. Diederich, "In Praise of Praise," NEA Journal, LII (September, 1963), 59.

"the error

feeling of

E.

comments m

in three s

basis of p

prepared a

teachers

groups:

except th

comments

and the

comment

treatme

on the

the "F

"Speci

group

He co

(Bos

Per

"the error avoiding approach has hardly given students a feeling of confidence and success."[67]

E. B. Page tested the effects of specified kinds of comments made by secondary teachers of various subjects in three school districts. Students were ranked on the basis of performance on teacher-made objective tests prepared as a part of the regular teaching plan. Then the teachers assigned them randomly to one of three treatment groups: the "No Comment" group, which received no marks except the grade; the "Free Comment" group, which received comments the teacher felt were appropriate for the student; and the "Specified Comment" group, which received uniform comments of an encouraging nature. The effects of the treatments were judged by examining the students' scores on the next objective test. Page found that in all schools the "Free Comment" group achieved higher scores than the "Specified Comment" group and that the "Specified Comment" group received better scores than the "No Comment" group. He concluded:

> When the average secondary teacher takes the time and trouble to write comments (believed to be "encouraging") on student papers, these apparently have a measurable and potent effect upon student effort, or attention, or attitude, or whatever it is which causes learning to improve, and this effect does not appear dependent on school building, school year or student ability.[68]

---

[67]James Moffett, Teaching the Universe of Discourse (Boston: Houghton Mifflin, 1968), p. 200.

[68]Ellis Batten Page, "Teacher Comments and Student Performance: A Seventy-Four Classroom Experiment in School

The study c

ments" were

that a stud

free comme

reasons fo

comments u

the effect

students

compositi

effects

the perf

students

essay t

or spec

student

period

studen

attitu

not fr

perfor

howeve

Motiva
(Augus

Attitu
Form O
of Wis

The study does not specify whether or not the "free comments" were generally positive or negative. It would seem that a study which provided an analysis of the nature of free comments would provide more information about the reasons for the positive effects of these kinds of comments upon students. One might speculate, also, about the effects of such comments upon the performance of students involved in the more complex process of written composition.

Roger Sweet conducted an experiment to test the effects of treatments like those specified by Page upon the performance and attitudes of ninth grade English students when the comment treatments were applied to four essay tests during a six-week period. He found that free or specified comments had minimal short-term effect on student performance on tests, but that over a longer period of time, free comments significantly increased student performance and positively affected student attitudes toward English.[69] The question of whether or not free comments alone caused students to improve their performance is not clearly answered by Sweet's study, however, since papers included letter grades as well as

---

Motivation," The Journal of Educational Psychology, XLIX (August, 1958), 180-81.

[69]Roger C. Sweet, "Educational Attainment and Attitudes Toward School as a Function of Feedback in the Form of Teachers' Written Comments" (Madison: University of Wisconsin, 1966), p. 10.

comments.

longer per

during whi

apparent r

study, rec

nature be

comments.

In

attitudes

expressed

students

treatment

essay; th

student's

grades we

papers ar

Resu
litt
lowe
fide
(3)
domi
writ

mentary

critici

Scrawl:
(ERIC
op. cit

comments. One must remember, also, that in this study, "a longer period of time" refers to the six-week period during which the experiment was conducted. Sweet, in apparent recognition of this weakness in the design of the study, recommended that a longitudinal study of a similar nature be conducted to further test the results of free comments.

In another study designed to test only student attitudes toward various combinations of reinforcement expressed in teachers' comments on essays, Clarke separated students in six eleventh grade English classes into nine treatment groups. The students then wrote an argumentative essay; the two teachers wrote comments appropriate to each student's treatment group on each of the papers, but no grades were assigned. Students read the comments on their papers and completed attitudinal questionnaires.

> Results showed that (1) the number of comments produced little effect; (2) purely negative comments produced lower scores in reinforcement, satisfaction, and confidence than completely positive comments produced; and (3) a mixture of criticism and praise, with praise dominating, produced the most satisfied and confident writers.[70]

Taylor and Hoedt, in a study conducted with elementary students, investigated the effects of praise and criticism upon fourth grade science students involved in a

---

[70]Grace Allison Clarke, "Interpreting the Penciled Scrawl: A Problem in Teacher Theme Evaluation," 1969 (ERIC No. ED 039 241), summarized in Daniel J. Dieterich, op. cit., pp. 1268-69.

creative
two rando
comments
whether t
elements
second gr
praise.
overlooke
two grou
comments
differer
of the
signifi
[and] w
those

the re
teach
The i
Illir
nega

Effe
Wri
(Oc

creative writing project. They divided the students into two random groups. One group of students received written comments of praise in response to their creative themes, whether they earned them or not. Teachers circled the best elements of these papers and indicated no errors. The second group received comments which lacked the element of praise. Teachers circled all errors on these papers and overlooked the best parts of them. Comments used for the two groups were carefully matched, and the number of comments controlled. The investigators found no significant difference in the quality of the creative writing endeavors of the two groups, but "those in the praised group produced significantly more work, exhibited more favorable attitudes, [and] were more highly motivated and more independent than those denied praise.[71]

One of the hypotheses in Biberstein's study tested the relationship between the evaluation procedures of the teacher and student performance in written composition. The investigation included fourth grade students in an Illinois school district. Biberstine concluded that negative marks by teachers on students' papers had a

---

[71]Winnifred F. Taylor and Kenneth C. Hoedt, "The Effect of Praise Upon the Quality and Quantity of Creative Writing," The Journal of Educational Research, LX (October, 1966), 83.

negative

comments

I

student w

effects c

students'

period, s

Students

an exper

ment.

assignm

essays

treatm

praise

on pap

sugges

of th

group

gesti

the p

turn

were

writ

Areas

Four

of I

negative effect on writing performance and that positive comments lost value when surrounded by negative ones.[72]

In yet another study of the effects of praise upon student writing, Thomas Gee collected data based upon the effects of comments he wrote in response to eleventh-grade students' expositiory compositions. During a four-week period, students wrote four essays on different topics. Students were aware of the fact that they were involved in an experiment but did not know the nature of the experiment. No instruction in composition skills related to the assignments was given. The investigator read each set of essays and applied three different types of comment treatments to the three groups of students: criticism, praise, and no comment. In addition, he marked all errors on papers of students in the criticized group and made suggestions for improvement but ignored the good aspects of the papers. He praised the good aspects of the praised group's papers, ignored their errors, and made no suggestions for improvements. He placed only a check mark on the papers of students in the no comment group. He returned each set of papers to students a week after they were written, and students read the comments before writing the next essay. Gee sought to determine the effects

---

[72]Richard Doyle Biberstine, "A Study of Selected Areas of Teacher Influence upon the Written Composition of Fourth Graders" (unpublished Ed.D. dissertation, University of Illinois, 1966), p. 113.

of the di

student w

sition.

pre- and

assessed

papers we

quality,

attitude

is affec

use; th

student

of tea

critic

attit

as a

of t

regu

One

ide

on

una

ad

va

Exp

"O

Mi

of the differing treatments upon quantity and quality of student writing and upon student attitudes toward composition. The first and fourth compositions were used as pre- and post-test measures. Quantity of writing was assessed by counting the number of T-units in a paper, papers were rated by three raters in order to determine quality, and a questionnaire was used to determine student attitudes. He concluded that quantity of student writing is affected by the type of written comment a teacher may use; that over a short period of time, the quality of student writing is not significantly influenced by the type of teacher comment; and that praise is more effective than criticism or no comment in developing positive student attitudes toward composition.[73]

It is unfortunate that Gee's study was conducted as an "experiment" unrelated to the learning activities of the class and with an investigator, rather than the regular classroom teacher, writing comments on the essays. One wonders if the outcomes obtained would have been identical if the regular teacher had written the comments on the students' papers, or if the students had been unaware of their involvement in an experiment. One advantage of Gee's study, however, is that there were few variables other than the differing comment treatments.

---

[73]Thomas Gee, "The Effects of Written Comment on Expository Composition" (unpublished Ph.D. dissertation, North Texas State University, 1970; Ann Arbor: University Microfilms, No. 71-551), pp. 65-66.

I

relative

ment, the

students'

of teache

comment"

scores on

study, an

marks by

student w

have beer

reinforce

papers di

quality c

writing p

gated the

as relate

Hoedt, ar

write mo

negative

studies

wonders

positive

have als

writing

In the studies which have thus far been reviewed relative to the effects of differing types of reinforcement, the findings have demonstrated improvement of students' attitudes when praise dominated in the responses of teachers. When achievement is considered, the "free comment" group in Page's investigation achieved better scores on objective tests than did the other groups in his study, and Biberstine's study demonstrated that negative marks by teachers on student papers had negative effects on student writing performance. In all other studies which have been reviewed, however, the different kinds of reinforcement teachers used in written comments on student papers did not lead to measurable differences in either quality of student writing performance or growth in student writing performance. In the two studies which investigated the effects of differing teacher comment treatments as related to the quantity of student writing (Taylor and Hoedt, and Gee), comments of praise caused students to write more than did students who received no comments or negative comments. It is unfortunate, however, that these studies were conducted over short periods of time. One wonders whether or not, over a longer period of time, the positive attitudes of students which were reported might have also caused significant improvement in student writing performance.

F

of praise

attitudes

teacher

praise w

successf

the most

the teac

importan

cation.

increase

not affe

seem to

Teaching

ideas wh

positive

instead,

ideas ar

p. 39.

Correct

Simon,
Merrill

Findings derived from the studies on the effects of praise do, however, support the previously cited attitudes of Paul Diederich toward the use of praise in teacher comments.[74] Dusel also recommended the use of praise when he advised teachers to show appreciation of successful writing. He went further to state that "one of the most effective . . . forms of appreciative response is the teacher's amplification of the pupil's idea."[75]

Student attitudes toward learning are regarded as important when one considers the traditional goals of education. Thus, if positive reinforcement does no more than increase the number of positive feelings of students while not affecting the quality of student learning, teachers seem to be justified in the extensive use of praise.

Raths, Harmin, and Simon, in their book Values and Teaching, have presented a method of responding to student ideas which probably should not be classified as either positive or negative reinforcement. The clarifying response, instead, is designed to help the student to consider his ideas and behavior and thus to clarify his values.[76]

---

[74]Diederich, "In Praise of Praise," op. cit., p. 39.

[75]Dusel, "Some Semantic Implications of Theme Correction," op. cit., p. 45.

[76]Louis E. Raths, Merrill Harmin, and Sidney B. Simon, Values and Teaching (Columbus, Ohio: Charles E. Merrill Publishing Company, 1966), p. 51.

T

several s

process (

element) v

of studen

as a grou

value-cla

certain p

found tha

One of th

est to te

clarifica

did not u

judged th

without k

Ratings c

which va

significa

class wh

two cours

impossib

process

performa

The authors of Values and Teaching summarized several studies in which the use of the value-clarification process (of which the clarifying response is a major element) was used in various settings with differing types of students. They stated that these investigations, taken as a group, contribute "some support for the assertion that value-clarification processes do make a difference in certain patterns of student behavior. In general, it was found that students became more purposeful and active."[77] One of the studies summarized in the book may be of interest to teachers of composition. Harry Shields used value-clarification processes in one of his college classes and did not use them in another. Three independent raters judged the quality of post-test papers from both classes without knowing which group had produced which papers. Ratings of the papers produced by students in the class in which value-clarification processes had been used were significantly better than the ratings of papers from the class which had been used as a control group. Since the two courses used by Shields differed in nature, it is impossible to be certain that the value-clarification process accounted for the differences in student writing performance which were reported in his study.[78]

---

[77] Ibid., p. 219.

[78] Ibid., p. 217.

S

clarifica

similar t

reinforce

alternati

is a tech

definitiv

of its ef

I

alternati

respondi

grading,

Numerous

fact tha

success

in varia

which he

theme re

English

justifie

his stud

English

for lay-

Studies which involved the use of value-clarification processes produced effects in students similar to the effects derived from the use of positive reinforcement. The use of the clarifying response as an alternative to either positive or negative reinforcement is a technique which should be investigated further before definitive statements can be made concerning the parameters of its effects upon student performance.

Alternatives to Written Teacher Responses on Student Essays

In recent years, teachers have explored various alternative or supplementary evaluative procedures for responding to student writing--lay readers, computer grading, taped oral commentary, and peer evaluation. Numerous articles in professional journals attest to the fact that these methods have met with varying levels of success and acceptance.

Evaluations of lay-reader programs have resulted in variant findings. Richard Clopper conducted a study in which he sought to determine whether or not the use of theme readers contributes to improvement in students' English skills and whether or not educational differences justified the expense of such a program. The readers in his study did not contribute to significant improvement in English skills; thus, he recommended that funds appropriated for lay-reader programs might better be used to reduce

class loa

State Dep

showed th

compared

classes.[8]

that "a r

more scho

it."[81]

A

the eleme

grammed

anything

and patie

will be

pointed

concerne

as a Mea
Skills"
Marylanc
of Inves
Secondar
1968), 7

Treatmer
Growth i
Educatio

Schools

Composi

class load.[79] In another study conducted by the Hawaii State Department of Education, however, lay reader classes showed the highest increase in essay rating scores when compared with tape evaluated, regular, and small group classes.[80] In a 1966 publication, Richard Corbin stated that "a recent study of the lay reader plan reports that more schools are abandoning the device than are adopting it."[81]

As for computer grading, Arthur Daigon described the elements of composition which a computer can be programmed to correct and observed that "a human being can do anything a computer can do if he has unlimited time, energy and patience."[82] Very soon, Daigon reported, the computer will be able to make comments to students, although he pointed out that the computer cannot honestly make comments concerned with personal response to student ideas. In

---

[79]Richard Clopper, "A Study of Contract Correcting as a Means of Significantly Increasing Writing and English Skills" (unpublished Ph.D. dissertation, University of Maryland, 1964), summarized in Nathan S. Blount, "Summary of Investigations Relating to the English Language Arts in Secondary Education: 1967," English Journal, LVII (May, 1968), 713-14.

[80]Patsy S. Saiki, "Comparison of the Lay Reader Treatment With Other Treatments in Increasing Student Growth in Writing, 1968-1969" (Hawaii: Department of Education, 1970), p. 31.

[81]Richard Corbin, The Teaching of Writing in Our Schools (New York: The Macmillan Company, 1966), p. 91.

[82]Arthur Daigon, "Computer Grading of English Composition," English Journal, LV (January, 1966), 52.

reviews

of essay

limitati

that the

Melville

resist d

as human

grading

structio

because

in an es

movement

Dartmouth

the compu

themes a

probably

of evalu

of Essay
of Educa
Educatio

English,

Company,

reviews of the Page and Paulus[83] study of computer analysis of essays, both Don Coombs and Ken Macrorie spoke to the limitations of computer evaluation. Coombs pointed out that the computer would give Hemingway low marks, and Melville high ones. He recommended: ". . . we should resist deciding that mechanized essay grading is as good as human . . . until we first consider whether human essay grading has demonstrable value in the process of instruction."[84] Macrorie criticized computer grading because the computer cannot respond to the quality of voice in an essay, one of the thrusts of the "New English" movement in the United States and Britain since the 1966 Dartmouth Conference.[85] In Uptaught, Macrorie stated that the computer could "analyze and rate a thousand student themes as well as a teacher because the themes were probably all dead."[86]

Many teachers have experimented with the dictation of evaluative comments into a dictaphone or tape recorder.

---

[83]Ellis B. Page and Dieter H. Paulus, "The Analysis of Essays by Computer, Final Report," United States Office of Education Project No. 6-1318 (Storrs, Conn.: Bureau of Educational Research, University of Connecticut, 1968).

[84]"Roundtable Review," Research in the Teaching of English, III (Fall, 1969), 228.

[85]Ibid., p. 233.

[86]Ken Macrorie, Uptaught (New York: Hayden Book Company, 1970), p. 6.

Stephen
essay ev
than the
students
United S
study at
use of t
ment of
was meas
teachers
four cla
were ide
comments
teacher
so that
written
spent ir
papers i
cost of
of the

Approac
1971),

Effecti
Teacher
sition,
Project

Stephen Vogler expressed enthusiasm for the tape method of essay evaluation. He stated that it takes no more time than the traditional method of evaluation and that his students responded positively.[87] In a study funded by the United States Office of Education, Jean McGrew conducted a study at a Lincoln, Nebraska, high school to evaluate the use of the dictation machine in the evaluation and improvement of student writing. When growth in writing performance was measured, the data favored the group for which the teachers had used the dictaphone method in three out of four classes, but not by a significant amount.[88] Problems were identified in the use of this method. The oral comments of the teachers were to be transcribed; thus, the teacher typically took pains to speak as he would write, so that the resulting comments would be in acceptable written form. This often resulted in doubling the time spent in evaluation. Also, the necessity of evaluating papers in a quiet place proved to be a hindrance. The cost of the typing was $.32 per paper evaluated. In spite of the limitations and costs, McGrew recommended that a

---

[87]Stephen H. Vogler, "Grading Themes: A New Approach; A New Dimension," English Journal, LX (January, 1971), 71.

[88]Jean Be. McGrew, "An Experiment to Assess the Effectiveness of the Dictation Machine as an Aid to Teachers in Evaluation and Improvement of Student Composition, Final Report," United States Office of Education Project No. 9-F-002, 1969, p. 16.

similar exp

than a seme

Supp

either supp

essays has

and Applebe

experience

observed.[89]

classes inc

class size

favored wi

noted, how

characteri

a study c

evaluatio

on the re

peer meth

the teach

tuting m

Methods
Press, I
Schoer,

Correcti
Hawaii:
Teachers
p. 10.

similar experiment extending over a total year, rather than a semester, be conducted.

Support for the practice of using peer evaluation to either supplement or supplant written teacher comments on essays has met with differing degrees of success. Squire and Applebee noted, however, that this important learning experience was virtually neglected in the schools they observed.[89] Dora Smith used this practice in the large classes included in her investigation of the effects of class size and found that "the large group was slightly favored with respect to composition."[90] It should be noted, however, that her students were ninth-graders characterized as high ability students. Pierson conducted a study comparing the effects of peer evaluation and teacher evaluation of the essays of ninth grade students. Based on the results of the study, his conclusion was that "the peer method seems to lead to the same measured results as the teacher method and has the added advantage of constituting much less of an after-hours problem for teachers."[91]

---

[89]Squire and Applebee, op. cit., p. 135.

[90]Dora V. Smith, Class Size in High School English: Methods and Results (Minneapolis: University of Minnesota Press, 1931), summarized in Braddock, Lloyd-Jones, and Schoer, op. cit., p. 98.

[91]Howard Pierson, "Peer Correction vs. Teachers' Correction of Writing" (paper delivered in Honolulu, Hawaii: Annual Convention of the National Council of Teachers of English, November, 1967; ERIC No. ED 017502), p. 10.

Peer evaluation was included as one of the evaluative techniques investigated by Sutton and Allen at Stetson University. During a ten-week period, one group of college freshmen participated in writing practice designed for evaluation by the teacher, and a second group wrote for evaluation by peers. A third group did no practice writing, but evaluated the compositions of peers. The investigators concluded that "none of these treatments can be said to have influenced writing performance as measured by this experiment in any significant way."[92]

Researchers who have investigated the utilization of peer evaluation, while not finding conclusive evidence to recommend its exclusive use, indicate that further investigation into the use of peer evaluation as a supplement to, or a substitute for, teachers' written comments is warranted. Lay reader programs, computer grading, and taped oral commentary, however, have not fared as well when they have been evaluated. At least until more research is done with alternative approaches to teachers' written comments as an evaluative technique, most teachers are likely to continue writing comments on students' essays.

---

[92]Sutton and Allen, op. cit., p. 58.

A r
sition eval
rated relia
variables a
Con
indicate wh
emphasized
Most exper
elements c
ment, but
cal error
whether i
to selec
recent o
teachers
comment
conduct
specifi
studen
on the
ate es

of re
of pr
othe

Conclusions Drawn From the Literature

A review of the literature concerned with composition evaluation indicates that students' essays can be rated reliably if raters are carefully trained and as many variables as possible are controlled.

Conclusive evidence does not exist, however, to indicate which aspects of writing performance should be emphasized when the teacher responds to student writing. Most experts advise that teachers respond to the larger elements of composition, such as organization and development, but some would also recommend the marking of mechanical errors as well. Similarly, experts disagree about whether it is best to mark student essays intensively, or to select a few things in the essay to emphasize. The more recent opinions, however, tend to support the idea that teachers should select only a few aspects of an essay to comment upon at any one time. More research needs to be conducted on the effects of the teacher's emphasis on specific aspects of composition when he writes comments on student essays. Until then, teachers are likely to operate on the basis of their personal prejudices when they evaluate essays.

Most investigations of the effects of specific kinds of reinforcement have shown that although teacher responses of praise lead to more favorable attitudes in students than other types of reinforcement, the kind of reinforcement the

teacher use

students' s

It

what does r

a differen

research i

evidence f

evaluation

value of t

teacher uses does not necessarily influence the quality of students' skill development.

It seems as though most research studies indicate what does not make a difference rather than what does make a difference when teachers evaluate student writing. More research is definitely needed in order to find conclusive evidence for the value of specific practices in the evaluation of written composition, and perhaps for the value of teacher evaluative practices.

PR

Th
subjects,
methods em
for statis

Th
Lincoln, N
English co
Northeast
schools wi
separate s
population
predominan
city of Li
excess of

Al
school dis
retained m

CHAPTER III

PROCEDURES FOR OBTAINING AND TREATING DATA

This chapter contains a detailed description of the subjects, the procedures used for collecting data, the methods employed in analyzing data, and the methods used for statistical treatment of the data.

Sources of the Data

The study was conducted at Northeast High School in Lincoln, Nebraska, where the investigator was a teacher of English composition during the 1972-1973 school year. Northeast High School is the second oldest of four high schools within the city. Built originally to serve three separate small towns within the city limits, the school's population of approximately 1,700 students is drawn from a predominantly lower-middle and middle class area of the city of Lincoln, a state capital with a population in excess of 150,000.

Although the school has been a part of the city school district for approximately thirty years, it has retained much of the style and flavor of a small high

school. T
settings f
The emphas
with the n
University
or the two
Du
required t
semester c
of this r
of the te
the remai
among va
study, s

emphasi
juniors
in the
was re
prepar
studer
a ran
but m
some

duri

school. The school and churches in the area provide the settings for a large proportion of the students' activities. The emphasis of the curriculum is upon providing students with the necessary background to enable them to enter the University of Nebraska, the technical schools in the area, or the two small private colleges located in the city.

During their three years at Northeast, students are required to complete twenty hours of English (four one-semester courses which meet five days per week). Ten hours of this requirement is satisfied by successful completion of the tenth grade course in English. In order to complete the remaining ten hours of English, students may choose among various courses in literature, composition, media study, speech, journalism, and humanities.

Composition I is a five-hour semester course emphasizing basic writing skills. It is open to both juniors and seniors, although typically only seniors enroll in the class. Until the 1971-1972 school year, the course was required of all seniors not enrolled in a college-preparatory course in English literature. Since that time, students enrolled in the composition course have represented a range of varying abilities with respect to English skills, but most of the students have expressed the desire to seek some form of education beyond high school.

One composition class taught by the investigator during the first semester of the 1972-1973 school year was

chosen for
study was
experiment
weeks of t

A
study, si
the amoun
students
Because t
amount of
skills of
of compos
sitions
measures
decision
presente
and subj
and then
essays w
students

. .
matt
test
atte

Schoer,
National

chosen for use in this study. Permission to conduct the study was granted by the Lincoln Public Schools. The experimental period extended throughout the first seventeen weeks of the semester.

Procedures for Collecting Data

A pretest-post-test design was chosen for this study, since the investigator was interested in measuring the amount of growth in writing performance of individual students included in two different treatment groups. Because the element of major concern in the study was the amount of growth in the total, general, and mechanical skills of written composition, the use of objective tests of composition skills was ruled out, and written compositions were used as the only pretest and post-test measures. The investigator found justification for this decision in Research in Written Composition. Braddock presented an analysis of several studies comparing objective and subjective tests of performance in written composition and then strongly recommended the use of several student essays when one wishes to evaluate the performance of students after a program of instruction:

> . . . if the teacher has been emphasizing rhetorical matters as well as grammar and mechanics, objective tests simply are not valid--they make little or no attempt to measure the larger elements of composition.[1]

[1]Richard Braddock, Richard Lloyd-Jones, and Lowell Schoer, Research in Written Composition (Champaign, Ill.: National Council of Teachers of English, 1963), p. 45.

gator
thirty
two gr
on all
the thi
eightee
two gro
initial
second
dropped
ferred t
December
an exten
assignme
from the
in the st
fifteen i

A
students
essays on
discourse
pretest a
Appendix
ments for
was based
written b

During the first week of the semester, the investigator used random selection procedures to divide the thirty-two students enrolled in the experimental class into two groups. One group would receive clarifying responses on all essays; the second group, directive responses. Of the thirty-two students enrolled in the experimental class, eighteen were girls and fourteen were boys. Each of the two groups formed by the use of random selection procedures initially contained nine girls and seven boys. During the second week of school, two girls in the clarifying group dropped the class; one girl in the directive group transferred to another section of Composition I. During December, one boy in the clarifying group was absent for an extended time and fell behind in the completion of his assignments; because of this, the investigator dropped him from the study. Thus, twenty-eight students were included in the study: thirteen in the clarifying group and fifteen in the directive group.

As a pretest measure of writing performance, students wrote in class, on three different days, three essays on topics representing three different modes of discourse: narration, exposition, and argumentation. The pretest assignments given to students are included in Appendix B of this study. The use of three writing assignments for pretest measurement of student writing ability was based on the finding of Kincaid that: "A single paper written by a student on a <u>given topic</u> at a <u>particular time</u>

cannot be considered as a valid basis for evaluating his achievement in a writing course at any time."[2] Kincaid concluded that the writing performance of poor writers varied according to the composition topic assigned; in an analysis of Kincaid's study, Braddock noted that differences existed not only in the topics assigned but also in the modes of discourse represented. Based on his analysis of Kincaid's findings, Braddock suggested that until more research has been done exploring the effect of mode of discourse on writing performance, this element should be controlled when planning assignments for an investigation such as the one undertaken in this study.[3]

When composing assignments for the three pretest essays, the investigator considered typical experiences and interests of high school students. She attempted to avoid topics which would provide advantages for any student or group of students. A study conducted by Bell in 1971 explored the written composition interests of high school students.[4] Topics he found to be of common interest to both boys and girls were "Education," "Our Society," "Life," "Our World," and "Experiences." Since Bell's topics do not qualify as good composition assignments when evaluated by the criteria recommended by the Commission on English, the

---

[2] Ibid., p. 95.

[3] Ibid., p. 9.

[4] James Baird Bell, "A Study of the Written Composition Interests of Senior High School Students" (unpublished Ed.D. dissertation, University of Oregon, 1971).

investigator used Bell's topics as general areas from which to draw more specific composition assignments. Then, noting the guidelines of the Commission on English,[5] the investigator attempted to compose assignments which would aid the students in formulating a proposition, which included data for the students to consider, and which stipulated the audience the students should address.

The investigator carefully matched pretest assignments with post-test assignments. The number of essays written, the examination situation, and the modes of discourse were the same for both tests, and the topics assigned were similar. (Pretest and post-test assignments are included in Appendix B.)

Students wrote the pretest essays during fifty-five minute class sessions. Braddock[6] advises that senior high school students should have seventy to ninety minutes in which to write a composition; since it was impossible to provide this amount of time in one block, the investigator decided to have students write in one fifty-five minute time block rather than to split the writing of a composition into two class sessions. Each examination situation was controlled, since all students wrote each pretest essay during the same period of the day.

---

[5]Commission on English, Freedom and Discipline in English (New York: College Entrance Examination Board, 1965), pp. 92-96.

[6]Braddock, Lloyd-Jones, and Schoer, op. cit., p. 9.

In order to ensure the anonymity of students, the investigator assigned each student a code number. Since the raters for the study were in no way associated with Northeast High School and the students in the experimental class, the investigator arranged the students' names in alphabetical order and assigned code numbers accordingly. Complete code numbers identified each student's treatment group, and a number which differed for each assignment was attached to the student code number. It was also important to conceal from the raters any indication of the time during the semester when any given set of essays included in the study had been written. The investigator photocopied all essays within two days after they were written, wrote code numbers on the back of the copies, sealed the copies in brown envelopes and stored them in a dark place for the duration of the experimental period.

After each set of essays was completed, the investigator, using the evaluation procedure described below, evaluated the essays. She wrote directive responses for all papers of students in the group designated to receive directive responses and clarifying responses for all papers of the students in the group designated to receive clarifying responses. In addition, at least one comment was written for each paper which identified something worthy of praise in the essay. The random sample of clarifying and directive responses included in Appendix C provide examples of both kinds of responses. All comments

were written on a separate sheet of paper, so that the investigator could begin evaluating the papers before she photocopied them. Typically, the teacher would write some of the comments in the margins of the essay. The examples in Appendix C, since they constitute a random sample drawn from all responses the teacher wrote during the experimental period, represent the typical length of responses and aspects of general writing performance to which the teacher responded. As can be seen, the investigator included comments related to ideas, organization, wording, and flavor in both the clarifying and directive responses. Also included in Appendix C are examples of student essays with clarifying and directive teacher responses.

Grades were not given on any of the essays written during the experimental period; the investigator did, however, record Diederich scale ratings for each essay on a score sheet she prepared for each student. Upon request, students could look at these ratings. Markings or comments related to student performance in the mechanical aspects of composition, such as usage, punctuation and spelling, were not placed upon the essays written during the experimental period. Instead, the teacher used a mechanics problems chart which she prepared for each student to indicate the frequency of specific mechanical errors. These charts provided baseline data for assigning mechanics problems assignments to each student. (A copy of the Mechanics Problems Chart is included in Appendix D.)

In order to provide for maximum student exposure to the directive or clarifying responses at the beginning of the experimental period, the students revised the three pretest essays, using one class period during the third week of the semester to revise each essay. Revision of these pretest essays early in the semester provided the teacher with a basis for observing the kinds of corrections students made in essays they had written rapidly during a limited examination period. Thus, the pretest essays and their revisions provided the teacher with more information about students' entering writing performance than the pretests alone would have provided. The concentration of revision activities, however, proved to be a tedious activity for many students; and teachers might well be able to gain sufficient information about a student's writing performance if they allowed him to choose one essay from the pretests for revision. After the pretest essays had been revised, students received the mechanics problems assignments. Each student attacked one mechanics problem at a time, beginning with basic problems and progressing through more complex problems. Each mechanics problems assignment included a statement of the problem to be attacked and provided a listing of reading assignments concerned with the specific problem. Each student was asked to study the assignments and to identify and correct related errors in the essays in his file folder. If he wished, he could seek help from the teacher or from fellow students who had

demonstrated competence in that specific skill. Students worked on mechanics problems periodically throughout the semester. Each student decided for himself how much time he wished to devote to this aspect of composition. When a student believed he had completed a problem satisfactorily, he submitted his work to the teacher and arranged for a conference with her. If the teacher found the student's work satisfactory, she gave him another mechanics assignment, with the provision that further work on the initial problem could be assigned if subsequent essays written by the student showed little or no improvement in the initial problem area.

The objectives for the Composition I course, of which one class was used as the experimental class in this study, are drawn from a number of sources. Muller, in writing about the 1966 Dartmouth Conference, stated that the seminar was interested in the value of writing for the development of children. "By writing they learn how to order and shape their experience, thereby learning more about life and themselves."[7] Hopefully, through the process of exploring his experience, each student should recognize that his perceptions of life have value; and through his involvement in the effective communication of these perceptions, his sense of self-worth should grow.

---

[7]Herbert J. Muller, The Uses of English (New York: Holt, Rinehart, and Winston, 1967), p. 98.

Fichetnau identified seven desirable outcomes of a composition program which are assumed to be representative of the stated objectives of many composition teachers. These outcomes include the student's ability to order ideas and show relationships between them; to think critically about the treatment of content; to see the relationship between purpose and form in written composition; to draw valid inferences and conclusions from data; to write coherent sentences, paragraphs, and sequences of paragraphs; and to follow the accepted conventions of grammar and usage.[8]

The author of this study accepted both the generally stated objective derived from the Dartmouth Conference and Fichtenau's objectives. The two sources are complementary, and the successful attainment of the Dartmouth objective is assumed to logically presuppose learning or student growth in one or more of the objectives stated by Fichtenau.

During the first few days of class, orientation to the procedures used in the class was interspersed with the writing and revision of the pretest essays. (A copy of the written course outline the teacher gave to students is included in Appendix E.) The teacher explained the objectives of the course and the content of the units of

---

[8]Robert L. Fichtenau, "Some Rhetorical Considerations for Teaching the Young Writer," English Journal, LIV (November, 1965), 720.

study. She did not tell the students that they were involved in an experiment. The investigator used the same basic procedures in all four sections of Composition I; these were the only sections of Composition I offered during the semester in which the experiment was conducted. Thus, students in the experimental class were not aware of differences they received in their class as compared with the treatment which students in the other Composition I classes received. (While the type of response used in the other classes was not controlled for experimental purposes, essays prepared by students in the non-experimental Composition I classes were rated by multiple raters. This procedure was possible because of the assistance of two student teachers who worked with the investigator in the three non-experimental classes.)

Grades in the high schools of Lincoln are issued in accordance with a seven point scale, with "1" being excellent and "7," failing. Students may also elect to register for a proportion of their classes on a pass-fail evaluation system. Grades are reported at two points in the semester: at the end of nine weeks and at the end of the semester. The investigator received permission to substitute either a "satisfactory" or "unsatisfactory" evaluation for a number grade at the nine-weeks reporting time, since nine-weeks grades are not recorded on the students' permanent records.

The investigator told her students that they would receive no grades on the individual essays they wrote for the class, but that they would receive written comments designed to help them improve their writing skills. She also told them that they were free to revise their essays as many times as they wished; and that if they wished to revise their essays beyond the first required revision, they could seek oral feedback from the teacher or from other students in the class. Students could see the numerical ratings for their essays if they wished, but no grades were associated with these ratings until the end of the semester, after the experiment was completed. Except for written responses on the papers and oral responses to individual students when initiated by the student, the teacher provided no feedback for the student between the time each essay was first turned in and the time the first revision was completed. Thus, the written responses of the teacher provided the major source of feedback to the student during the time between completion of the rough draft and completion of the first revision for each essay.

The teacher presented to her students a copy of the Diederich scale which would be used for the rating of their essays and explained the eight factors within it. Students were encouraged to attempt to rate their own compositions using this scale, or to ask other students to rate their essays according to the scale and to explain their reasons for rating each factor as they did.

The investigator told students they would earn a grade of at least "6" or "pass" for the semester if they completed all of the work assigned during the semester, but that they could not earn a passing grade unless they completed all of the assigned work. She also informed them that their grade for the semester would be based upon their scores on the post-test essays written at the end of the semester and upon evidence of significant improvement during the semester. She explained the use of the "satisfactory-unsatisfactory" evaluation for nine-weeks reporting. All students who had completed work assigned through the first nine weeks of the semester would receive a "satisfactory" evaluation. Also, students and their parents would receive a written evaluation which described their progress in the course and identified their major strengths and weaknesses. This procedure was followed for all sections of Composition I taught by the investigator.

The Composition Units

During the experimental period, instruction in written composition centered on five units of study:

1. Personal Narration
2. Description
3. Definition
4. Exposition
5. Argumentation

Each unit included instruction in three stages of the writing experience: prewriting, writing, and evaluation-revision. The teacher explained the major writing assignment to students on the first day of each unit. Thus, the essays that resulted from each unit of instruction differed from the pretest and post-test essays in that they were formulated over a period of time during which students participated in prewriting experiences. The investigator found support for this procedure in the writings of experts in the teaching of composition and in an experiment entitled The Composing Processes of Twelfth Graders, by Janet Emig. Professor Emig analyzed the composing processes of eight sixteen- and seventeen-year-old secondary school students. In her study she observed that writers rarely proceed through the writing process without interruption.[9] She noted that, when students engage in self-sponsored writing, prewriting is a much longer process than in school-sponsored writing, in which there is often no time provided for prewriting.[10]

Prewriting experiences for each unit consisted of instructional techniques designed to help students discover various approaches to the types of writing called for in

---

[9]Janet Emig, The Composing Processes of Twelfth Graders (Urbana, Ill.: National Council of Teachers of English, 1971), p. 40.

[10]Ibid., p. 92.

the units. These experiences included common stimuli for the entire class (such as films, classroom discussions, and assigned readings) which would help students formulate ideas for their written compositions and small-group discussions in which students could present their ideas to their peers for clarification. Often, students brought short practice paragraphs or essays to these small group sessions for evaluation.

Each student brought to class a rough draft of the major writing assignment for each unit on a designated day at the end of the prewriting activities. Students met in small groups to read and evaluate their papers. Throughout the units, the teacher provided situations in which students could interact with their classmates and discover the principles of effective composition. Professor Emig noted in her study that "the most significant others in the private, and often the school sponsored, writing of twelfth-graders are peers. . . ."[11] Thus, in the current study, the investigator attempted to use these "most significant others" as much as possible in her instructional approaches. The use of small group evaluation provided each student with an audience for his writing and allowed for feedback from fellow students.

---

[11] Ibid., p. 100.

On the day following small group evaluation, the students completed and submitted the first drafts of their essays to the teacher for evaluation. The teacher then photocopied these essays for later use by the raters. The original papers were then evaluated by the investigator, directive or clarifying responses were written, and the papers were returned to students for revision.

In an article about the evaluation of written composition, Paul Diederich commented: "The average English paper corrected by the average English teacher looks as though it had been trampled on with cleated boots and has about the same effect on the student."[12] The objective of the investigator in developing and using the approach described was to eliminate some of this "trampling" through the use of an individualized and sequential program for modification of the writing problems of her students. The teacher's comments on papers of students in both the directive and clarifying groups emphasized the development and organization of ideas and aspects of wording and flavor; the mechanics chart provided a record of the mechanical writing problems of students. She did not pressure students to work on mechanical elements of writing, but as the semester progressed, many students showed an interest in improving this aspect of their writing so that

---

[12]Paul B. Diederich, "In Praise of Praise," NEA Journal, LII (September, 1963), 58.

they might express their ideas more forcefully. The teacher asked students to consider her written comments and to make whatever revisions they thought necessary. Because many unit essays were too long to revise in one class period, the teacher provided two class periods for revision, collecting the essays at the end of the first period and returning them at the beginning of the second. The teacher photocopied the revised essays and stored the copies for later use by the raters. Teacher response to the revised essays was brief, usually indicating an area of improvement.

On the day following the first revision, students worked individually on assignments related to mechanical problems. On the second day following the revision, the revised papers were returned to the students, and students met in small groups to read their papers. At times, the instructor read several essays to the class. Some students wished to make further improvements in their essays after the unit learning activities had been completed. The investigator encouraged students to revise their essays as many times as they wished. Students completed all revisions beyond the first revision outside of class.

The Rating of the Essays

At the close of the experimental period the following essays were rated by a team of raters: the three pretest essays and the first revisions of them, the five unit essays and the first revisions of them, and the three

post-test essays. In all, the data consisted of nineteen sets of essays from the experimental class.

Three persons were chosen to rate these essays. Two of the raters hold doctorates in education and professional rank in both the department of Secondary Education and the department of English at the University of Nebraska-Lincoln. The third rater, chairman of the English department of another high school in the city, has had experience rating College Board essay examinations. Two of the raters rated all of the essays included in the study, and the third rater rated only those essays for which the ratings of the first two raters were discrepant by five or more points.[13] The third rater rated these essays after the first two raters had completed their ratings.

Prior to the rating of the essays, the investigator conducted a training session for the raters. During the first meeting, she explained the Diederich rating scale which she had chosen as the instrument to be used for rating the essays in the study. (A copy of the scale is

[13]This procedure is recommended by Paul B. Diederich and Frances R. Link, "Cooperative Evaluation in English," in Evaluation as Feedback and Guide, ed. by Fred T. Wilhelms (Washington, D.C.: Association for Supervision and Curriculum Development, NEA, 1967), p. 201. Diederich suggests that a third rater rate those essays for which the first two ratings are discrepant by ten points; to increase the reliability of the ratings, the investigator submitted all papers for which the first two ratings were discrepant by more than five points to a third rater.

included in Appendix A). After discussing the definitions of the various factors in the scale, the raters rated four of the essays from a practice set of thirty essays the investigator had prepared for rater training. The essays used for rater training were selected from essays prepared by students in the nonexperimental classes of Composition I. The raters compared their rating scores and discussed their reasons for rating the papers as they had. They then took the remaining twenty-six essays with them and rated them during the following week. When these ratings were returned, the investigator found only four essays for which the three ratings were discrepant by more than ten points. After using the Pearson product moment correlation coefficient to compute inter-rater reliability, the investigator found that the results approached significance for the number of essays in the practice set used for training. She decided, however, that the reliability of the ratings would be increased if the margin of allowable inter-rater discrepancy were reduced to five points.

A second training session was held with the raters so that they might discuss the rating scale further and re-rate the essays for which their ratings had been discrepant by more than ten points. At the conclusion of the second training session, the two primary raters were provided with their initial sets of papers for rating.

In preparing the sets of essays for rating, the investigator decided to keep all essays related to one

assignment together. Since the essays represented first drafts and revisions of many different assignments written in various modes of discourse, this procedure allowed the raters to experience an evaluation situation similar to the situation a classroom teacher encounters when he evaluates a set of papers. As they rated the papers, the raters were able to establish mental comparisons between the essays which would aid them in establishing a consistent base for their ratings. It was important, however, that the raters not know the order in which the essays were written and that they be unable to distinguish first draft essays from revisions and pretest essays from post-test essays. The coding system used to identify assignments, described earlier in this chapter, obliterated the sequence of assignments. In addition, the investigator arranged the sets of essays in a different random order for each of the two major raters. Thus, the two initial raters rated the nineteen sets of essays in a differing sequence. In order to guard against effects of rater fatigue which might influence the ratings of individual essays within each set of essays, the investigator arranged the essays within each package in a different random order for each of the two major raters.

The investigator delivered two sets of essays to each major rater each week during the rating period and picked up these sets at the end of the week. Although it

would have been desirable for the raters to evaluate the essays in the same place during the same time, the number of essays in the study and the demands upon the raters' time made this arrangement possible. The two raters did, however, evaluate all of the essays during the same nine-week period, at the rate of two sets of essays per week during the first eight weeks and three sets during the ninth week. The investigator gave the raters the following typewritten instructions prior to the rating of the essays:

1. The papers in each package are arranged in the order in which you should rate them. Do not deviate from this order.

2. Read the Diederich Scale before rating each set of papers.

3. Before rating any papers in a set, choose four papers at random and scan them. After scanning, be certain to return these papers to their original order in the package before beginning to rate.

4. Return to the first paper in the set and begin rating. Rate each paper as rapidly as possible. You should spend no more than three minutes on a typical paper. Do not write on the papers as you are rating them. After you have read the paper, circle the appropriate numbers on the rating card and total the score for the paper. Initial the rating card; then record the code number which appears on the back of the last page of the essay. (Do not look at the code number until you have finished rating the essay.)

5. If you rate both sets of essays on the same day, rest at least thirty minutes after rating the first set before beginning the second set.

6. After you have finished rating a package of essays, set aside the essays and do not look at them again.

After the two major raters had finished rating the essays, the investigator recorded the scores. She then

selected those essays for which the two ratings were discrepant by five or more points and delivered these to the third rater for rating in packages of approximately thirty papers. The two scores which were closest were then chosen as the ratings for these essays. If the three ratings were equidistant, the two lower ratings were chosen.

A mean score was computed for each essay by averaging the two rating scores. Pretest scores for each student were computed by averaging the three mean scores from the three pretest essays. Post-test scores were computed using this same procedure.

The investigator wished to determine the extent to which her written comments for the first draft essays in the study complied with the definitions she had established for clarifying and directive responses. Thus, after the raters had completed the rating of all essays included in the study, she gave each of the raters definitions of clarifying and directive responses and copies of thirty-five sets of written comments randomly selected from the comments she had written in response to the 224 first draft essays included in the study. She asked the raters to determine the type of response each of the thirty-five sets represented. The results of this validation of responses is reported in Chapter IV.

Analysis of the Data

Levels of inter-rater reliability and the internal reliability of the ratings used in the study were computed. The levels of inter-rater reliability achieved by the use of the Pearson product moment correlation coefficient (r_{xy})[14] are reported in Chapter IV.

The data obtained from the ratings of all nineteen examples of each student's writing performance were organized into tables which provide summarizations for individuals and for both treatment groups. From the tabled data, the planned statistical treatments were then conducted for each of the stated hypotheses of the study. Tabled data appear in Chapter IV.

Treatment of the Data

An alpha level of .05 was established as the level of statistical significance required for the rejection of the hypotheses of this study. Since all hypotheses are stated in the null form, rejection of any hypotheses permits the acceptance of the alternative hypothesis, i.e., differences do exist between treatment groups. Use of an .05 level of significance as the criterion reduces the probability of an error in judgment to a level no greater than one instance from each twenty observations. When the null

---

[14]William L. Hays, Statistics (New York: Holt, Rinehart and Winston, 1963), pp. 497-98.

hypothesis cannot be rejected, however, it should be noted that this does not necessarily mean that there are no differences between the variables; rather, the failure to reject the null hypothesis can only be interpreted as indicating that the presence of differences has not been established so as to permit conclusions about the relationship between the variables being studied.

All three research hypotheses are concerned with changes in student performance in three categories: total performance, general performance, and mechanical skills performance. For each of the major hypotheses, the relationships within each of these categories were statistically treated by use of a t-test. Basic information was transferred from the summary sheets and was coded on computer programming sheets. Computer cards were prepared and the t-tests required were computed through the use of the computer facilities of the University of Nebraska.

CHAPTER IV

PRESENTATION AND ANALYSIS OF THE DATA

The basic purpose of this study was to determine whether or not the investigator's use of two different types of written comments in response to student essays produced significant differences in the amount of growth in writing performance of two groups of twelfth grade students. The three hypotheses of this study permit the reporting of findings comparing differences in growth in total, general, and mechanical writing performance for the two groups of students (1) over the period of time for which the experiment extended (one academic semester), (2) during the period of time between the initial essays students wrote in response to one unit of study and the initial essays students wrote in response to a contiguous unit of study, and (3) between initial drafts and revisions of eight sets of essays (three pretest essays and five unit essays).

The Diederich scale used for rating the essays in the study provided for the reporting of scores related to specific aspects of general, mechanical, and total writing performance. Both the directive and clarifying comments

which the teacher wrote centered on aspects of general writing performance; thus, data related to general writing performance are of major interest in the study. Data related to total writing performance were included to permit the identification of instances in which significant changes in general writing performance were of a magnitude to produce significant differences in overall writing performance. Data related to mechanical writing performance were included for two reasons: (1) to ascertain whether or not significant differences in general writing performance influenced significant differences in mechanical writing performance, and (2) to provide basic information about changes in mechanical writing performance occurring when no comments regarding mechanical aspects were included in the evaluative responses to essays of students in either of the two experimental groups.

Reliability of the Rating Procedures Used

Included in Table 1 are data which were obtained from the computation of both inter-rater reliability and internal reliability of the ratings used in the study. In computation of the internal reliability of ratings actually used in the study, when ratings by the two major raters were discrepant by five or more points, these ratings were compared with the rating by the third rater, and the two

Table 1.--Inter-Rater Reliability and Reliability of the Ratings Used for Data Analysis: Expressed in Pearson Product-Moment Correlation Coefficients (r).

Reliability Data and Findings	Rater A Compared to Rater B	Rater A Compared to Rater C	Rater B Compared to Rater C	Ratings Used in Study
N	532	213	213	532
Sum (X)	18,626	7,447	7,088	18,265
Sum (Y)	18,203	7,069	7,069	18,183
Sum (X^2)	663,775	265,335	255,834	642,725
Sum (Y^2)	649,999	241,994	241,994	635,886
Sum (XY)	644,083	248,780	240,370	636,487
r	.38	.27	.42	.81
S_{r_o}	.0434	.0687	.0687	.0434
z-score*	8.76	3.93	6.11	18.66
Significance	.0001	.0001	.0001	.0001

*A z-score (standard score), obtained by dividing the Pearson product-moment correlation coefficient by the standard error of the Pearson product-moment correlation coefficient (S_{r_o}), is significant at the .0001 level of significance when it equals or exceeds 3.70.

ratings with the least discrepancy were chosen. If the three ratings were equidistant, the two lower ratings were chosen.

The procedure the investigator used for computation of the Pearson product-moment correlation coefficient (r) is described in Basic Statistical Methods.[1] The formulas used for the computation of the standard error of the Pearson product-moment correlation coefficient (S_{r_o}) and the standard score (z-score) are also described in Basic Statistical Methods.[2] Table 2 of the same text was used to determine the level of significance of the obtained z-scores.[3] This procedure is preferable to the simple reporting of obtained correlations, a common practice in many experimental studies reviewed in Chapter II of this study.

Validation of Teacher Written Responses Used in the Study

Since the basic problem in this study involved the investigation of the effects of two differing styles of teacher written responses used on student essays, the investigator implemented a procedure to test and validate the assumption that the comments used in the study actually represented two distinct types of comments.

---

[1] N. M. Downie and R. W. Heath, Basic Statistical Methods (New York: Harper and Row, 1965), p. 85.

[2] Ibid., pp. 155-56.

[3] Ibid., pp. 290-97.

After the raters had rated all papers, the investigator provided each of the raters with a random sample of thirty-five sets of the comments which she had written in response to the 224 first draft essays which were included in the study. She provided the raters with definitions of the two types of responses and asked them to identify each response as either a directive response or a clarifying response. The set of directions and the form the raters used to code the responses according to type is reproduced in Appendix C. The resulting data are summarized in Table 2, which indicates the extent to which the raters correctly identified the type of response the teacher had used. Each of the thirty-five sets of responses was correctly identified by at least two of the raters, and twenty-nine of the thirty-five responses were correctly identified by all three raters. The obtained chi-square value for the data presented in Table 2 was 83.06. Table IV in Basic Statistical Methods indicates that a chi-square value of 13.815 is required for the .001 level of significance; higher levels are not provided in the table.[4] The obtained value is far in excess of the level required for significance at the .001 level; thus, the null hypothesis used for validation--that no significant difference existed between the types of comment and that raters would assign comments to categories on a random basis--was

[4] Ibid., p. 299.

Table 2.--Correctness of Rater Identification of Type of Written Response Used by the Investigator in Responding to Student Essays.

Rater	Correctly Identified Type of Response Used	Incorrectly Identified Type of Response Used
A	32	3
B	35	0
C	32	3
Totals	99	6

rejected. The data indicated that there were distinctive differences in the style of the two types of comments used as written teacher responses to the student essays used as sources of the data.

Analysis of the Data of the Study

With the use of the computer facilities at the University of Nebraska, student t-tests were computed for the three null hypotheses in this investigation. As indicated in the design of the study, an alpha level of .05 had been established as the level of statistical significance required for the rejection of the three hypotheses. This criterion was used in interpretation of the data.

Intergroup Comparison of Pretest and Post-test Writing Performance

The primary hypothesis investigated in this study was related to the differences in long-term growth in

writing performance between students in the group which received clarifying responses and students in the group which received directive responses. The first hypothesis, stated in the null form, was:

> When pretest and post-test scores of total, general, or mechanical writing performance, as measured by the Diederich composition rating scale, are compared by the use of a student t-test, there is no significant difference between the growth in writing performance of the group of students who received clarifying responses for their written compositions and the growth in writing performance of the group of students who received directive responses for their written compositions.

Analysis of the data related to the testing of the first hypothesis yielded the findings which are summarized in Table 3.

Since membership in the two groups was determined at the beginning of the experiment by random assignment, normal assumptive procedures would permit the conclusion that no significant differences existed between the two groups at the beginning of the experiment. The t-values reported for the pretest essays, as shown in Table 3, confirm the conclusion that the two groups were not significantly different at the beginning of the experiment in general, mechanical, or total writing performance. The reported t-values for the post-test essays indicate that there were also no significant differences in general, mechanical, or total writing performance at the end of the experiment. When growth scores in writing performance during the experiment were compared, the obtained t-values

Table 3.--Analysis of Intergroup Comparisons of General, Mechanical, and Total Writing Performance for Pretest Scores, Post-test Scores, and Growth in Writing Performance During the Experimental Period.

Area of Writing Performance	Clarifying Group Mean (N = 13)	Directive Group Mean (N = 15)	Level of Significance (t-value)
General			
Pretest	19.46	18.31	.09
Post-test	20.02	19.41	.29
Growth	.56	1.10	.30
Growth (as %)	2.88	6.01	. .
Mechanical			
Pretest	14.01	14.07	.47
Post-test	14.81	14.79	.48
Growth	.80	.72	.42
Growth (as %)	5.71	5.12	. .
Total			
Pretest	33.47	32.38	.21
Post-test	34.83	34.20	.36
Growth	1.36	1.82	.36
Growth (as %)	4.06	5.62	. .

indicated no significant differences in growth between the clarifying group and the directive group in any of the three categories--general, mechanical, or total writing performance. No statistically significant differences were found in the analysis of the data for hypothesis I; thus, it was accepted.

The growth scores in Table 3 are also reported in percentage form. As was stated in Chapter II of this study, Braddock noted that annual gains in writing performance are usually reported as about 5 per cent.[5] Since the experiment described in this study extended for only a semester, the gains reported in this study compare favorably with the average annual gains noted by Braddock.

Intergroup Comparisons of Growth in Writing Performance Between Contiguous Units of Study

The second of the three major hypotheses investigated in this study was concerned with the differences in growth in writing performance between units of study for students in the clarifying group as compared with students in the directive group. The second hypothesis, stated in the null form, was:

> When scores of total, general, or mechanical writing performance on initial compositions for one unit of study, as measured by the Diederich composition rating

---

[5]Richard Braddock, Richard Lloyd-Jones, and Lowell Schoer, Research in Written Composition (Champaign, Ill.: National Council of Teachers of English, 1963), p. 7.

> scale, are compared by use of a student t-test with scores of total, general, or mechanical writing performance on initial compositions for a contiguous unit of study, there is no significant difference between the growth in writing performance of the group of students who received clarifying responses for their written compositions and the growth in writing performance of the group of students who received directive responses for their written compositions.

Analysis of the data related to the testing of the second hypothesis yielded the findings which are summarized in Tables 4, 5, and 6. Table 4 summarizes the findings related to growth in general writing performance between the two groups; Table 5, the findings related to growth in mechanical writing performance between the two groups; and Table 6, the findings related to the growth in total writing performance between the two groups.

In Table 4, the obtained t-values comparing the mean scores of the two groups for each of the five unit assignments indicated no significant differences in the general writing performance of the two groups of students on the first drafts of each unit assignment. Thus, significant differences which occurred in growth scores between units probably resulted from the varying teacher responses. The mean scores reported in Table 4 indicate that both the clarifying group and the directive group declined in general writing performance between the writing of the initial draft of the essay for Unit I (Personal Narration) and the initial draft of the essay for Unit II (Description). The t-value of .33, however, indicates that there were

Table 4.--Analysis of Intergroup Comparisons of Growth in General Writing Performance Occurring Between Contiguous Units of Study.

Unit: Contiguous Unit	Clarifying Group Mean (N = 13)	Directive Group Mean (N = 15)	t-value
Unit #1: Unit #2			
Unit #1: Personal Narration	22.15	21.10	.22
Unit #2: Description	21.19	18.53	.12
Growth Score	-.96	-2.57	.33
Unit #2: Unit #3			
Unit #2: Description	21.19	18.53	.12
Unit #3: Definition	18.00	19.47	.15
Growth Score	-3.19	.94	.03*
Unit #3: Unit #4			
Unit #3: Definition	18.00	19.47	.15
Unit #4: Exposition	20.38	21.37	.20
Growth Score	2.38	1.90	.37
Unit #4: Unit #5			
Unit #4: Exposition	20.38	21.37	.20
Unit #5: Argumentation	20.96	18.93	.06
Growth Score	.58	-2.44	.02*

*Level of significance for significant findings.

no significant differences between the two groups in the amount of decline shown.

When the initial drafts of essays students wrote for Unit II (Description) are compared with the initial drafts of essays students wrote for Unit III (Definition), the mean scores indicate that the Clarifying group declined 3.19 points in general writing performance as measured by the Diederich scale, and the Directive group increased .94 in general writing performance. The t-value of .03 indicates that there was a significant difference in the amounts of change in general writing performance for the two groups between Unit II and Unit III, and that this difference favored the Directive group. The only instructional variable during this time was the type of teacher response to descriptive essays.

When the mean scores reported for Unit III (Definition) are compared with the mean scores for Unit IV (Exposition), the resulting growth scores indicate that both the Clarifying and Directive groups grew in general writing performance. The reported t-value indicates that there were no significant differences between the two groups in the amount of growth shown.

Comparison of the mean scores for Unit IV (Exposition) with the mean scores for Unit V (Argumentation) indicates that the Clarifying group grew in general writing performance by .58 and the Directive group declined

in general writing performance by 2.44 points. The resulting t-value of .02 identifies a significant difference between the two groups in the amount of growth in general writing performance; the difference favors the Clarifying group. The instructional variable during this period was the type of response to students' expository essays.

The findings summarized in Table 5 are related to the amounts of growth in mechanical writing performance between contiguous units of study for the two groups of students. A comparison of the mean scores reported shows two instances of decline for the Clarifying group and one instance of decline for the Directive group. In all other instances, the table shows slight gains in growth in mechanical writing skills for both groups of students. The reported t-values in Table 5 indicate no significant differences in the amounts of growth in mechanical writing performance shown by the Clarifying and Directive groups. Also, there were no significant differences in the mechanical writing performance of the two groups on the initial drafts of each of the unit assignments.

The findings summarized in Table 6 show comparisons of contiguous essays in the amounts of growth in total writing performance (the combination of general and mechanical writing performance) between the two groups of students. In Table 6, the obtained t-values comparing

Table 5.--Analysis of Intergroup Comparisons of Growth in Mechanical Writing Performance Occurring Between Contiguous Units of Study.

Unit: Contiguous Unit	Clarifying Group Mean (N = 13)	Directive Group Mean (N = 15)	t-value
Unit #1: Unit #2			
Unit #1: Personal Narration	15.93	14.70	.12
Unit #2: Description	14.93	15.67	.31
Growth Score	-1.00	.97	.16
Unit #2: Unit #3			
Unit #2: Description	14.93	15.67	.31
Unit #3: Definition	15.35	14.63	.41
Growth Score	.42	-1.04	.46
Unit #3: Unit #4			
Unit #3: Definition	15.35	14.63	.41
Unit #4: Exposition	15.27	14.70	.40
Growth Score	-.08	.07	.48
Unit #4: Unit #5			
Unit #4: Exposition	15.27	14.70	.40
Unit #5: Argumentation	15.42	15.10	.31
Growth Score	.15	.40	.44

Table 6.--Analysis of Intergroup Comparisons of Growth in Total Writing Performance Occurring Between Contiguous Units of Study.

Unit: Contiguous Unit	Clarifying Group Mean (N = 13)	Directive Group Mean (N = 15)	t-value
Unit #1: Unit #2			
Unit #1: Personal Narration	38.08	35.80	.14
Unit #2: Description	36.12	34.20	.16
Growth Score	-1.96	-1.60	.50
Unit #2: Unit #3			
Unit #2: Description	36.12	34.20	.16
Unit #3: Definition	33.35	34.10	.37
Growth Score	-2.77	-.10	.10
Unit #3: Unit #4			
Unit #3: Definition	33.35	34.10	.37
Unit #4: Exposition	35.65	36.07	.42
Growth Score	2.30	1.97	.43
Unit #4: Unit #5			
Unit #4: Exposition	35.65	36.07	.42
Unit #5: Argumentation	36.38	34.03	.09
Growth Score	.73	-2.04	.05*

*Level of Significance for significant findings.

the mean scores of the two groups for each of the five unit assignments indicated no significant differences in the total writing performance of the two groups of students on the first drafts of each unit assignment. Table 6 indicates only one area of significant difference in growth in total writing performance when the two groups of students are compared. A comparison of the mean growth scores reported for total writing performance on the intial drafts of Unit IV (Exposition) and Unit V (Argumentation) indicates that the Clarifying group grew .73 and the Directive group declined by 2.04 points. The reported t-value of .05 indicates that between Units IV and V there was a significant difference between the amount of growth in total writing performance shown by the Clarifying group and the amount of growth shown by the Directive group. This difference favored the Clarifying group. This is an instance wherein a significant difference in general writing performance produced a significant difference in total writing performance.

When all findings related to the second hypothesis, summarized in Tables 4, 5, and 6, were considered, no significant differences existed between the two groups of students in general, mechanical, or total writing performance on the first drafts of any of the five unit assignments. Significant differences were found between the amounts of growth in general writing performance of

the two groups of students for two of the four time spans which were investigated. The difference for the period between the description and definition units favored the Directive group, and the difference for the time between the expository and argumentation units favored the Clarifying group. No significant differences in the amounts of growth in mechanical writing performance existed for any of the time spans investigated. A significant difference, favoring the Clarifying group, was shown to exist between the amounts of growth in total writing performance for the two groups for the time between the exposition and argumentation units. Because the findings showed significant differences in the amounts of growth for the two groups in three comparisons of interest in hypothesis II, the null hypothesis was rejected for these three instances. For the remaining nine comparisons of interest in hypothesis II, however, no significant differences in the amounts of growth were found to exist between the two groups; thus, the null hypothesis was accepted for these nine instances.

Intergroup Comparisons of Growth in Writing Performance between First Draft Essays and Revisions

The third, and final, major hypothesis investigated in this study was related to the differences in amounts of growth in writing performance for the Clarifying and

Directive groups of students between first draft essays and their revisions. The third hypothesis, stated in the null form, was:

> When the scores of total, general, or mechanical writing performance on the initial drafts of written compositions, as measured by the Diederich composition rating scale, are compared by use of a student t-test with the scores of total, general, or mechanical writing performance on the revisions of these compositions, there are no significant differences between the amount of change in these scores for the group of students who received clarifying responses for their written compositions and the group of students who received directive responses for their written compositions.

Analysis of the data related to the testing of the third hypothesis yielded the findings which are summarized in Tables 7, 8, and 9. Table 7 summarizes the findings related to growth in general writing performance between the two groups of students; Table 8, the findings related to growth in mechanical writing performance between the two groups; and Table 9, the findings related to the growth in total writing performance between the two groups.

Comparisons of the mean scores for the first draft essays and their revisions, as reported in Table 7, indicate significant differences between the general writing performance of the Clarifying and Directive groups for two essays. In both cases these differences existed in revisions. The reported t-value of .04 indicates that the two groups differed significantly in their general writing performance on the revision of Pretest I (Personal Narration) and on the revision of the Unit IV essay

Table 7.--Analysis of Intergroup Comparisons of Growth in General Writing Performance Occurring as a Result of the Revision Process.

Number and Title of Pretest or Unit Assignment	Sources of Data Used	Clarifying Group Mean (N = 13)	Directive Group Mean (N = 15)	t-value
Pretest #1: Personal Narration	First Draft	19.50	17.78	.13
	Revision	20.08	17.93	.04*
	Growth Score	.58	.15	.38
Pretest #2: Exposition	First Draft	17.88	17.60	.38
	Revision	20.42	18.77	.06
	Growth Score	2.54	1.17	.09
Pretest #3: Argumentation	First Draft	21.00	19.50	.09
	Revision	19.00	17.60	.14
	Growth Score	-2.00	-1.90	.47
Unit #1: Personal Narration	First Draft	22.15	21.10	.22
	Revision	21.38	18.70	.07
	Growth Score	-.77	-2.40	.30
Unit #2: Description	First Draft	21.19	18.53	.12
	Revision	21.04	20.27	.24
	Growth Score	-.15	-1.74	.26
Unit #3: Definition	First Draft	18.00	19.47	.15
	Revision	18.96	18.33	.37
	Growth Score	.96	-1.14	.16
Unit #4: Exposition	First Draft	20.38	21.37	.20
	Revision	20.65	18.63	.04*
	Growth Score	.27	-2.74	.01*
Unit #5: Argumentation	First Draft	20.96	18.93	.06
	Revision	21.32	19.27	.08
	Growth Score	.36	.34	.47
Summary:	Net Growth	.23	-.60	.11

*Level of significance for significant findings.

(Exposition). In both instances where significant differences were found, the differences favored the Clarifying group. When changes in the amounts of growth were considered, the Clarifying group declined slightly in general writing performance on three of the revised essays and grew slightly on five of the revised essays. The Directive group declined slightly in general writing performance on five of the revised essays and grew slightly in their revisions of three essays.

When growth scores were analyzed, only one area of significant difference between the two groups existed. When mean scores for the first drafts and revisions of Unit IV (Exposition) essays were compared, the data showed a mean gain of .27 for the Clarifying group and a mean decline of 2.74 points for the Directive group. The reported t-value of .01 indicated a significant difference in the amounts of growth in general writing performance on the expository revision for the two groups of students, favoring the Clarifying group.

When net growth scores for general writing performance were compared, the two groups showed no significant difference in mean growth scores for revisions.

The mean scores for mechanical writing performance on first drafts and revisions reported in Table 8 indicate that the Clarifying group declined in mechanical writing performance on four of the eight revisions and improved on four of the eight revisions. The Directive group declined

Table 8.--Analysis of Intergroup Comparisons of Growth in Mechanical Writing Performance Occurring as a Result of the Revision Process.

Number and Title of Pretest or Unit Assignment	Sources of Data Used	Clarifying Group Mean (N = 13)	Directive Group Mean (N = 15)	t-value
Pretest #1:	First Draft	13.65	13.72	.50
Personal Narration	Revision	14.80	13.57	.17
	Growth Score	1.15	-.15	.11
Pretest #2:	First Draft	13.54	13.70	.40
Exposition	Revision	15.12	14.46	.18
	Growth Score	1.58	.76	.13
Pretest #3:	First Draft	14.85	14.83	.50
Argumentation	Revision	14.73	14.53	.40
	Growth Score	-.12	-.30	.36
Unit #1:	First Draft	15.93	14.70	.12
Personal Narration	Revision	14.04	14.50	.28
	Growth Score	-1.89	-.20	.24
Unit #2:	First Draft	14.93	15.67	.31
Description	Revision	15.00	13.99	.16
	Growth Score	.07	-1.68	.26
Unit #3:	First Draft	15.35	14.63	.41
Definition	Revision	14.50	15.20	.34
	Growth Score	-.85	.57	.42
Unit #4:	First Draft	15.27	14.70	.40
Exposition	Revision	14.97	13.91	.12
	Growth Score	-.30	-.79	.12
Unit #5:	First Draft	15.42	15.10	.31
Argumentation	Revision	15.60	15.20	.20
	Growth Score	.18	.10	.35
Summary:	Net Growth	-.03	-.21	.07

in mechanical writing performance on five of the revisions and improved on three revisions. When the mean growth scores for the two groups were compared, no significant differences were found to exist in growth in mechanical writing performance on any of the eight revisions. Mean net growth scores for mechanical writing performance indicated that both groups declined slightly. When these net growth scores were compared, no significant difference was found to exist between the two groups.

Comparisons of the mean scores for total writing performance on first drafts and revisions reported in Table 9 indicate significant differences between the total writing performance of the Clarifying and Directive groups for two essays. In both cases these differences existed in revisions. The reported t-value of .04 indicates that the two groups differed signficantly in their total writing performance on the revision of Pretest I (Personal Narration). The second t-value of .04 indicates that the two groups also differed significantly in their total writing performance on the revision of the Unit IV essay (Exposition). When changes in amounts of growth were considered, the Clarifying group declined slightly in total writing performance on four of the revised essays and improved slightly on four of the revised essays. The Directive group also declined slightly in total writing

Table 9.--Analysis of Intergroup Comparisons of Growth in Total Writing Performance Occurring as a Result of the Revision Process.

Number and Title of Pretest or Unit Assignment	Sources of Data Used	Clarifying Group Mean (N = 13)	Directive Group Mean (N = 15)	t-value
Pretest #1: Personal Narration	First Draft	33.15	31.50	.23
	Revision	34.88	31.50	.04*
	Growth Score	1.73	.00	.16
Pretest #2: Exposition	First Draft	31.42	31.30	.47
	Revision	35.54	33.23	.09
	Growth Score	4.12	1.93	.08
Pretest #3: Argumentation	First Draft	35.85	34.33	.18
	Revision	33.73	32.13	.20
	Growth Score	-2.12	-2.20	.48
Unit #1: Personal Narration	First Draft	38.08	35.80	.14
	Revision	35.42	33.20	.12
	Growth Score	-2.66	-2.60	.29
Unit #2: Description	First Draft	36.12	34.20	.16
	Revision	36.04	34.26	.18
	Growth Score	-.08	.06	.47
Unit #3: Definition	First Draft	33.35	34.10	.37
	Revision	33.46	33.53	.49
	Growth Score	.11	-.57	.35
Unit #4: Exposition	First Draft	35.65	36.07	.42
	Revision	35.62	32.54	.04*
	Growth Score	-.03	-3.53	.01*
Unit #5:	First Draft	36.38	34.03	.09
	Revision	36.92	34.47	.09
	Growth Score	.54	.44	.48
Summary	Net Growth	.20	-.81	.05*

*Level of significance for significant findings.

performance on four revisions, grew slightly on three revisions, and did not change on one revision.

When growth scores for individual sets of essays were analyzed, only one area of significant difference between the two groups existed. When mean scores for the first drafts and revisions of Unit IV (Exposition) essays were compared, the data showed a mean decline of .03 for the Clarifying group and a mean decline of 3.53 for the Directive group. The reported t-value of .01 indicated a significant difference in the amount of decline in total writing performance on the expository revision for the two groups of students, with the Clarifying group declining less than the Directive group. (It should be noted that the net decline of .03 for the Clarifying group resulted from a decline in mechanical writing performance, since the Clarifying group grew .27 in general writing performance.)

Net growth scores for total writing performance on revisions indicated an increase of .20 for the Clarifying group and a decline of .81 for the Directive group. The reported t-value of .05 indicated that a significant difference existed between the growth in total writing performance of the two groups on revision. This difference favored the Clarifying group.

When the total findings related to the twenty-seven comparisons of interest for hypothesis III--summarized in Tables 7, 8, and 9--were considered, significant differences

were found to exist between the amounts of growth in writing performance which occurred as a result of revision for the two groups in three instances. The Clarifying group grew significantly more in general writing skills on the Unit IV (Exposition) revision than did the Directive group (which declined). The Directive group declined significantly more than the Clarifying group in total writing skills on the Unit IV revision. The Clarifying group grew significantly more in net growth for total writing performance than did the Directive group, which declined. Because of these three significant findings, the third hypothesis was rejected for these three comparisons. For the remaining comparisons, however, no significant differences between the two groups of students were found to exist, and the null hypothesis was accepted for these twenty-four comparisons.

Summary of the Findings

In summary, the findings related to the three comparisons of interest for hypothesis I showed that no significant differences in growth in general, mechanical, or total writing performance existed between the Clarifying and Directive groups at the end of the experimental period. The null form of the first hypothesis was accepted totally.

The findings related to the twelve comparisons of interest for hypothesis II indicated significant differences in the amounts of growth in writing performance for

the two groups between contiguous units of study for three of the twelve comparisons. Two of three differences existed in general writing performance; one favoring the Directive group for the period between the description and definition units; the other, favoring the Clarifying group for the time between the exposition and argumentation units. The third difference existed in total writing performance for the time between the exposition and argumentation units and favored the Clarifying group. The null hypothesis was rejected for these three comparisons and accepted for the remaining nine comparisons.

The findings related to the twenty-seven comparisons of interest for hypothesis III indicated significant differences between the two groups in the amount of growth which resulted from revisions in three comparisons. Two of the significant differences between the two groups existed in the general and total writing performance for the revision of the expository essay; and in both cases, the differences favored the Clarifying group. The third significant difference between the two groups in growth which resulted from revision existed in the net growth scores for total writing performance. This difference again favored the Clariflying group. The third hypothesis, stated in the null form, was rejected for the three comparisons described above and accepted for the twenty-four remaining comparisons. A summary of the findings

Table 10.--A Summary of the Findings of the Study.

Hypotheses	Number of Comparisons	Significant Findings		Nonsignificant Findings
		C	D	
Hypothesis #1				
General	1			1
Mechanical	1			1
Total	1			1
Hypothesis #2				
General	4	1	1	2
Mechanical	4			4
Total	4	1		3
Hypothesis #3				
General	8	1		7
Mechanical	8			8
Total	8	1		7
Net Growth Score	3	1		2
Summary:				
General	13	2	1	10
Mechanical	13			13
Total	13	2		11
Net Growth Score	3	1		2
Totals	42	5	1	36

related to the three major hypotheses of this study is provided in Table 10.

CHAPTER V

SUMMARY, CONCLUSIONS, IMPLICATIONS AND RECOMMENDATIONS

The purpose of this experimental study was to investigate the effects of two types of written teacher response to student essays upon students' growth in general, total and mechanical writing performance. The two types of responses were defined as directive and clarifying. The study sought to determine the effects of these types of response upon student growth in the three categories of writing performance over a substantial period of time, between units of study, and between first-drafts and revisions of essays. This chapter includes a brief summary of the methods and procedures used in the experiment and a summary of the findings of the study. In addition, conclusions drawn from the data and implications and recommendations for further investigations are presented.

Summary of Methods and Procedures

The subjects for this experiment were twenty-eight twelfth grade students in a one-semester composition class which the author taught. Students in the class were divided randomly into two groups; one group of students received clarifying responses for essays they wrote during the semester, and the other group received directive responses. The type of response the student received was the only variable in this experiment. Students wrote three pretest essays in response to assignments representing different modes of discourse as a measure of entering writing performance and three similar post-test essays as a measure of ending writing performance. During the experimental period, students received written responses for the three pretest essays and five major unit essays and revised these eight essays. Thus, nineteen samples of writing for each of the twenty-eight students (532 essays in all) formed the data base for this experiment. At the end of the experimental period, a team of three raters rated the essays, using the Diederich composition rating scale. Ratings for these essays were used in computing student t-tests for the comparisons of interest in the three major hypotheses of this study. The findings of these statistical analyses were presented in Chapter IV.

Summary of the Findings

Each of the hypotheses was considered separately in Chapter IV, and tables were presented for the t-test analyses. Analysis of the data resulted in the following findings:

1. The findings related to the three comparisons of interest for the first hypothesis indicated that no significant differences in long term growth in general, mechanical, or total writing performance existed between the Clarifying group and the Directive group.

2. The findings related to the twelve comparisons of interest for the second hypothesis indicated two instances of significant growth in general writing performance between contiguous units of study--one favoring the Directive group for the period between the description and definition units, the other favoring the Clarifying group for the period between the exposition and argumentation units. No significant differences in mechanical writing performance between contiguous units existed. One instance of significant difference in growth in total writing performance favoring the Clarifying group existed for the time between the exposition and argumentation units. In the remaining nine comparisons of interest, no significant differences between the two groups existed.

3. The findings related to the twenty-seven comparisons of interest for the third hypothesis indicated three instances of significant differences in the growth in writing performance as a result of revision. These three differences favored the Clarifying group. One difference occurred in the growth in general writing performance for the expository revision and a second occurred in the growth in total writing performance for the expository revision. The third difference existed in the net growth score which resulted from the pooling of the growth scores for the eight revisions. In the remaining twenty-four comparisons of interest, no significant differences existed between the Clarifying and Directive groups.

Observation of the findings related to all hypotheses indicated six significant findings and thirty-six non-significant findings. When the total findings of the study are summarized according to the three categories of writing performance, it can be noted that no significant differences existed between the two groups for the thirteen comparisons related to growth in mechanical writing performance. Two significant differences favoring the Clarifying group and one significant difference favoring the Directive group existed for the thirteen comparisons related to growth in general writing performance. Two significant differences favoring the Clarifying group existed for the thirteen comparisons related to growth in total writing performance. One significant difference favoring the Clarifying group existed for the net growth in total writing performance occurring as a result of the revision process.

Conclusions

From the findings included in this study, the investigator concluded:

1. There is no assurance that a teacher's choice of either clarifying responses or directive responses to student essays produces an advantage which influences the amount of growth in total, general, or mechanical writing performance of twelfth grade students during the course of an academic semester.

2. There is no conclusive evidence that the use of either the clarifying or directive teacher response produces an advantage which consistently influences the amount of growth in writing performance which

occurs between contiguous units of study, although the clarifying response may be more effective than the directive response when used in an exposition unit immediately followed by an argumentation unit.

3. There is no conclusive evidence that a teacher's choice of either the clarifying or directive response on initial drafts of essays produces an advantage which consistently influences the amount of growth in writing performance of twelfth grade students on revisions, but the clarifying response may be more effective than the directive response for producing growth on revisions of expository essays.

4. There is no assurance, when a twelfth grade student revises an essay after the teacher has used the clarifying or directive response, that the revised essay will show evidence of better writing performance than that shown on an initial draft of an essay, although net growth scores for revisions over a total semester may show a slight advantage for the use of the clarifying response.

Implications and Recommendations

The findings of this study indicated no definite advantage for either of the two types of teacher response in producing growth in the writing performance of twelfth grade students over the period of one academic semester. The findings reported in Tables 4 and 7 of Chapter IV do, however, show that the Clarifying group improved consistently in general writing performance during the last three units of the experimental period, whereas the Directive group experienced a net decline during the last three units. Further experiments extending for at least an academic year should be conducted in order to determine the effects of the two types of comments over an extensive period. Also, since this experiment was limited by the

inclusion of students at only one grade level, the study should be replicated with students at various grade levels.

The findings of the study indicated that no advantage existed for either type of teacher response for influencing growth in mechanical writing performance; thus, neither type of response demonstrated an advantage for producing a "spillover" effect upon mechanical writing performance. Growth in mechanical writing skills was substantial for both groups, however; and this growth suggests that it may not be necessary for teachers to refer to mechanical aspects of writing in their written responses to essays. Experiments should be conducted in which the effects of directive or clarifying responses in combination with mechanical corrections are compared with the effects of directive or clarifying responses without mechanical corrections.

While responding to individual essays included in the study, the investigator sometimes believed that a directive response might have been more effective than a clarifying response (or vice versa) for a certain aspect of an essay. Future experiments might be designed to investigate the effects of a combination of directive and clarifying responses upon growth in writing performance.

The fact that several modes of discourse were included in the composition course used for this experiment complicates conclusions related to growth between units

during the experimental period. There is an indication that the use of clarifying responses for expository essays may produce more growth between consecutive expository and argumentative units than directive responses produce. There is also an indication that the use of directive responses may produce more growth in general writing performance between consecutive descriptive and definition units. (See Table 4 in the previous chapter.) Although significant findings support both of these indications, these significant findings related to growth between units resulted from the fact that one group declined sharply while the other group grew slightly in writing performance. More research related to various sequences of units within composition courses is needed before conclusive statements about the effects of clarifying or directive responses upon growth between units in writing courses can be made. Also, the effects of the use of directive and clarifying responses with several essays for each specific mode of discourse should be investigated.

The findings related to the effects of directive and clarifying responses upon revisions suggested that there was no consistent advantage for the use of either type of response with all kinds of essays included in this study. There is a strong indication, however, that the clarifying response may be more effective than the directive response for expository essays. Findings related to exposition reported for the second hypothesis also support this

conclusion. Further experiments should be designed to investigate the effects of directive and clarifying responses upon several samples of expository writing.

This study was not designed to test the effects of clarifying or directive comments upon individual students; however, one student in the clarifying group improved consistently on all revisions he wrote during the experimental period, and one student in the directive group gained significantly more in writing performance over the course of the semester than did her classmates. This suggests that future experiments designed to test the effects of clarifying and directive comments upon individual students and specific types of learners may be of value.

One of the limitations of this study was the use of only one teacher who wrote directive and clarifying responses for student essays. Future studies might investigate the effects of directive and clarifying comments upon student writing performance when teachers with various teaching styles use these comments. Perhaps student response to teacher comments may be related to ways in which students respond to a teacher's personality or teaching method.

Many revision scores for essays in the study were lower than scores for corresponding first drafts. Teachers have long assumed that revision of essays should be a

major composition learning activity. This study did not examine the revision process as an isolated variable, but the findings related to revision scores suggest that more research should be conducted to investigate changes in writing performance resulting from revision which comes after various kinds of evaluative procedures. Previous studies concerned with the effects of the revision process have neglected comparisons of first drafts and revisions. The way to discover what happens when students revise is to examine actual revisions. More research studies which include essays and revisions written during an extended experimental period are needed.

The author of this study cannot determine the proportion of the total growth in writing performance which should be attributed to revision activities, since there may have been learning from the revision process which caused later writing performance to improve. A comparison of net growth scores for revision with growth scores over the experimental period, however, would suggest that much of the learning which takes place in composition classes should be attributed to factors other than revision activities. An experiment should be designed to investigate the effects of clarifying and directive teacher responses when revision is not required. More importantly, many more experiments should be initiated to isolate the proportionate amounts of learning which take place in

various kinds of prewriting periods and rewriting periods. Further research with methods of prewriting instruction might indicate that teachers could stimulate effective growth in students' writing performance if they spent more time planning lessons and less time evaluating essays.

This study--like many others which have been reviewed--has contributed more to knowledge about what does not make a difference than what does make a difference when teachers evaluate student writing. A few years ago, West called for more valid research studies in composition and suggested that "research in composition remains in a kind of pre-scientific era."[1] The investigator concurs with West and recommends that more research in all areas of the teaching and learning of composition be conducted. More importantly, teachers must become skilled in research procedures, since teachers have the daily opportunity for learning how students improve their writing skills.

---

[1]William W. West, "Written Composition," Review of Educational Research, XXXVII (April, 1967), 167.

BIBLIOGRAPHY

BIBLIOGRAPHY

Books

Braddock, Richard; Lloyd-Jones, Richard; and Schoer, Lowell. Research in Written Composition. Champaign, Ill.: National Council of Teachers of English, 1963.

Commission on English. Freedom and Discipline in English. New York: College Entrance Examination Board, 1965.

Corbin, Richard. The Teaching of Writing in Our Schools. New York: The Macmillan Company, 1966.

Diederich, Paul B. "Grading and Measuring." Improving English Composition. Edited by Arno Jewett and Chalres Bish. Washington, D.C.: National Education Association, 1965.

_____; French, John W.; and Carlton, Sydell T. Factors in Judgments of Writing Ability. Research Bulletin 61-15. Princeton: Educational Testing Service, 1961.

_____, and Link, Frances R. "Cooperative Evaluation in English." Evaluation as Feedback and Guide. Edited by Fred T. Wilhelms. Washington, D.C.: Association for Supervision and Curriculum Development, National Education Association, 1967.

Downie, N. M., and Heath, R. W. Basic Statistical Methods. New York: Harper and Row, 1965.

Emig, Janet. The Composing Processes of Twelfth Graders: NCTE Research Report No. 13. Urbana, Ill.: National Council of Teachers of English, 1971.

Godshalk, Fred I.; Swineford, Francis; and Coffman, William E. The Measurement of Writing Ability. New York: College Entrance Examination Board, 1966.

Greene, Harry A. "English Language, Grammar, and Composition." Encyclopedia of Educational Research. Edited by W. S. Monroe. New York: Macmillan, 1950.

Hays, William L. Statistics. New York: Holt, Rinehart and Winston, 1963.

Hilgard, E. R., and Bower, G. H. Theories of Learning. New York: Appleton-Century-Crofts, 1966.

Kitzhaber, Albert. Themes, Theories, and Therapy: The Teaching of Writing in College. New York: McGraw-Hill Book Company, 1963.

LaBrant, Lou L. We Teach English. New York: Harcourt, Brace, 1951.

Macrorie, Ken. Uptaught. New York: Hayden Book Company, 1970.

Meckel, Henry C. "Research on Teaching Composition and Literature." Handbook of Research on Teaching. Edited by N. L. Gage. Chicago: Rand McNally and Company, 1963.

Moffett, James. Teaching the Universe of Discourse. Boston: Houghton, Mifflin, 1968.

Monroe, W. S. "The Unreliability of the Measurements of Ability in Written Composition." Twenty-second Yearbook of the National Society for the Study of Education, Part I. Bloomington: Public School Publishing Company, 1923.

Muller, Herbert J. The Uses of English. New York: Holt, Rinehart, and Winston, 1967.

Raths, Louis E.; Harmin, Merrill; and Simon, Sidney B. Values and Teaching. Columbus, Ohio: Charles E. Merrill Publishing Company, 1966.

Remmers, H. H., and Gage, N. L. Educational Measurement and Evaluation. New York: Harper and Brothers, 1943.

Sister M. Judine, I. H. M., ed. A Guide for Evaluating Student Composition. Champaign, Ill.: National Council of Teachers of English, 1965.

Smith, Dora V. Class Size in High School English: Methods and Results. Minneapolis: University of Minnesota Press, 1931.

Squire, James, and Applebee, Roger K. High School English Instruction Today. New York: Appleton-Century-Crofts, 1968.

Articles

Arnold, Lois V. "Writer's Cramp and Eyestrain--Are They Paying Off?" English Journal, LIII (January, 1964), 10-15.

Blount, Nathan S. "Summary of Investigations Relating to the English Language Arts in Secondary Education: 1967." English Journal, LVII (May, 1968), 710-24.

Briggs, Dennis. "The Influence of Handwriting on Assessment." Educational Research, XIII (November, 1970), 50-55.

Cast, B. M. D. "The Efficiency of Different Methods of Marking English Composition, Part II." British Journal of Educational Psychology, X (February, 1940), 49-60.

Christiansen, Mark. "Tripling Writing and Omitting Readings in Freshman English: An Experiment." College Composition and Communication, XVI (May, 1965), 122-24.

Coffman, William E. "On the Validity of Essay Tests of Achievement." Journal of Educational Measurement, III (1966), 151-56.

Confrey, A. "An Investigation of the Comments Made by Forty Instructors upon Students' Themes." Catholic Education Digest, XXIV (1927), 335-40.

"Corrections Hurt Writers, Expert Says." College Composition and Communication, XIV (May, 1963), 84.

Daigon, Arthur. "Computer Grading of English Composition." English Journal, LV (January, 1966), 46-52.

Denby, Robert V. "NCTE/ERIC Report--Composition Evaluation." English Journal, LVII (November, 1968), 1215-21.

Diederich, Paul B. "How to Measure Growth in Writing Ability." English Journal, LV (April, 1966), 435-49.

_____. "In Praise of Praise." NEA Journal, LII (September, 1963), 58-59.

Dieterich, Daniel J. "Composition Evaluation: Options and Advice." English Journal, LXI (November, 1972), 1264-71.

Dressel, Paul; Schmid, John, Jr.; and Kincaid, Gerald. "The Effect of Writing Frequency upon Essay-Type Writing Proficiency at the College Level." Journal of Educational Research, XLVI (December, 1952), 285-93.

Dusel, William J. "Determining an Efficient Teaching Load in English." Illinois English Bulletin, Vol. XLIII, No. 1 (1955), 44-48.

_____. "Some Semantic Implications of Theme Correction." English Journal, XLIV (March, 1955), 390-97.

Fellows, John Ernest. "The Influence of Theme-Reading and Theme Correction on Eliminating Technical Errors in the Written Compositions of Ninth Grade Pupils." University of Iowa Studies in Education, Vol. VII, No. 1 (1932), 7-44.

Fichtenau, Robert L. "Some Rhetorical Considerations for Teaching the Young Writer." English Journal, LIV (November, 1965), 720-23.

Follman, J. C.; Miller, W. G.; Lowe, A. J.; and Stefurak, D. W. "Effects of Time and Typeface on Level and Reliability of Theme Grades." Research in the Teaching of English, IV (Spring, 1970), 51-58.

Fostoedt, Donald. "Criteria for the Evaluation of High School Composition." Journal of Educational Research, LIX (November, 1965), 108-12.

Green, James L. "A Method for Writing Comments on Student Themes." English Journal, LVII (February, 1968), 215-22.

Heys, Frank, Jr. "The Theme-a-Week Assumption: A Report of an Experiment." English Journal, LI (May, 1962), 320-22.

Hillegas, M. B. "A Scale for the Measurement of Ability in English Composition by Young People." Teachers College Record, XII (September, 1912), 331-84.

Lokke, Virginia, and Wykoff, George. "Double Writing in Freshman Composition--an Experiment." School and Society, XLVIII (November, 1948), 437-39.

Page, Ellis Batten. "Teacher Comments and Student Performance: A Seventy-Four Classroom Experiment in School Motivation." The Journal of Educational Psychology, XLIX (August, 1958), 173-81.

"Roundtable Review." Research in the Teaching of English, III (Fall, 1969), 222-36.

Smith, Vernon H. "Measuring Teacher Judgment in the Evaluation of Written Composition." Research in the Teaching of English, III (Fall, 1969), 181-95.

Solomon, Daniel, and Rosenberg, Larry. "Teacher-Student Feedback and Classroom Social Structure." Journal of Social Psychology, LXII (April, 1964), 197-210.

Taylor, Winnifred F., and Hoedt, Kenneth C. "The Effect of Praise upon the Quality and Quantity of Creative Writing." The Journal of Educational Research, LX (October, 1966), 80-83.

Thorndike, E. L. "A Scale of Merit in English Writing by Young People." Journal of Educational Psychology, XI (June, 1911), 361-68.

Vernon, P. E., and Millican, G. D. "A Further Study of the Reliability of English Essays." British Journal of Psychology, VII (November, 1954), 65-76.

Vogler, Stephen H. "Grading Themes: A New Approach; A New Dimension." English Journal, LX (January, 1971), 70-74.

West, William W. "Written Composition." Review of Educational Research, XXXVII (April, 1967), 159-67.

Willing, M. H. "Measurement of Written Composition." English Journal, VII (March, 1918), 193-202.

Publications of Government Agencies

Burton, Dwight L., and Arnold, Lois V. "Effects of Frequency of Writing and Intensity of Evaluation upon High School Students' Performance in Written Composition." United States Office of Education Cooperative Research Project 1523. Tallahassee: Florida State University, 1963.

Clark, William G. "An Evaluation of Two Techniques of Teaching Freshman Composition: Final Report." Colorado Springs: Air Force Academy, 1968.

McGrew, Jean B. "An Experiment to Assess the Effectiveness of the Dictation Machine as an Aid to Teachers in Evaluation and Improvement of Student Composition, Final Report." United States Office of Education Project No. 9-F-002, 1969.

Page, Ellis B., and Paulus, Dieter H. "The Analysis of Essays by Computer, Final Report." United States Office of Education Project No. 6-1318. Storrs, Conn.: Bureau of Educational Research, University of Connecticut, 1968.

McColly, William. "The Dimensions of Composition Annotation." United States Office of Education Cooperative Research Project S-216. Oswego: State University of New York at Oswego, 1965.

_____, and Remstad, Robert. "Comparative Effectiveness of Composition Skills Learning Activities in the Secondary School." United States Office of Education Cooperative Research Project 1528. Madison: University of Wisconsin, 1963.

Report 3, 1969-1970 Writing: National Results. Ann Arbor: National Assessment of Educational Progress, 1970.

Saiki, Patsy S. "Comparison of the Lay Reader Treatment with Other Treatments in Increasing Student Growth in Writing, 1968-1969." Hawaii: Department of Education, 1970.

Sutton, Joseph T., and Allen, Eliot D. "The Effects of Practice and Evaluation on Improvement in Written Composition." United States Office of Education Cooperative Research Project No. 1993. Deland, Florida: Stetson University, 1964.

Unpublished Materials

Bell, James Baird. "A Study of the Written Composition Interests of Senior High School Students." Unpublished Ed.D. dissertation, University of Oregon, 1971.

Biberstine, Richard Doyle. "A Study of Selected Areas of Teacher Influence upon the Written Composition of Fourth Graders." Unpublished Ed.D. dissertation, University of Illinois, 1966.

Buxton, Earl W. "An Experiment to Test the Effects of Writing Frequency and Guided Practice upon Students' Skill in Written Expression." Unpublished Ph.D. dissertation, Stanford University, 1958.

Clarke, Grace Allison. "Interpreting the Penciled Scrawl: A Problem in Teacher Theme Evaluation." Educational Resources Information Center, No. ED 039 241, 1969.

Clopper, Richard. "A Study of Contract Correcting as a Means of Significantly Increasing Writing and English Skills." Unpublished Ph.D. dissertation, University of Maryland, 1964.

Gee, Thomas. "The Effects of Written Comment on Expository Compositions." Unpublished Ph.D. dissertation, North Texas State University, 1970.

Kincaid, Gerald L. "Some Factors Affecting Variations in the Quality of Students' Writing." Unpublished Ed.D. dissertation (Michigan State College), Michigan State University, 1953.

Maize, Ray C. "A Study of Two Methods of Teaching English to Retarded College Freshmen." Unpublished Ph.D. dissertation, Purdue University, 1952.

Marshall, Barbara Ruth Cohn. "A Survey and Analysis of Teachers' Markings on Selected Compositions of Average Students in Grades 10 and 12." Unpublished Ed.D. dissertation, State University of New York at Buffalo, 1971.

McMechan, Melville Y. "The Relative Effectiveness of Four Procedures for Evaluating Students' Written Themes." Unpublished M.A. thesis, University of British Columbia, 1961.

Pierson, Howard. "Peer Correction vs. Teachers' Correction of Writing." Paper delivered in Honolulu, Hawaii: Annual Convention of the National Council of Teachers of English, November, 1967. Educational Resources Information Center, No. ED 017 502.

Sweet, Roger C. "Educational Attainment and Attitudes toward School as a Function of Feedback in the Form of Teachers' Written Comments." Madison: University of Wisconsin, 1966. Educational Resources Information Center, No. ED 015 163.

Whalen, Thomas E. "A Validation of the Smith Test for Measuring Teacher Judgment of Written Composition." Paper presented at the annual meeting of the American Educational Research Association, New York, February, 1971. Educational Resources Information Center, No. ED 049 275.

APPENDICES

APPENDIX A

DIEDERICH SCALE

APPENDIX A

DIEDERICH SCALE*

Topic _____ Reader _____ Paper _____

	Low		Middle		High
Ideas	2	4	6	8	10
Organization	2	4	6	8	10
Wording	1	2	3	4	5
Flavor	1	2	3	4	5
Usage	1	2	3	4	5
Punctuation	1	2	3	4	5
Spelling	1	2	3	4	5
Handwriting	1	2	3	4	5

I. GENERAL MERIT

1. Ideas

High. The student has given some thought to the topic and has written what he really thinks. He discusses each main point long enough to show clearly what he means. He supports each main point with arguments, examples, or details; he gives the reader some reason for believing it. His points are clearly related to the topic and to the main idea or impression he is trying to get across. No necessary points are overlooked and there is no padding.

---

*This scale and its description are extracted from pp. 444-46 of "How to Measure Growth in Writing Ability," by Paul B. Diederich, English Journal, LV (April, 1966), 435-49.

Middle. The paper gives the impression that the student does not really believe what he is writing or does not fully realize what it means. He tries to guess what the teacher wants and writes what he thinks will get by. He does not explain his points very clearly or make them come alive to the reader. He writes what he thinks will sound good, not what he believes or knows.

Low. It is either hard to tell what points the student is trying to make or else they are so silly that he would have realized that they made no sense if he had only stopped to think. He is only trying to get something down on paper. He does not explain his points; he only writes them and then goes on to something else, or he repeats them in slightly different words. He does not bother to check his facts, and much of what he writes is obviously untrue. No one believes this sort of writing--not even the student who wrote it.

2. Organization

High. The paper starts at a good point, moves in a straight line, gets somewhere, and stops at a good point. The paper has a plan that the reader can follow; he is never in doubt as to where he is or where he is going. Sometimes there is a little twist near the end that makes the paper come out in a way that the reader does not expect, but it seems quite logical. Main points are treated at greatest length or with greatest emphasis; others in proportion to their importance.

Middle. The organization of this paper is standardized and conventional. There is usually a one-paragraph introduction, three main points each treated in one paragraph, and a conclusion that often seems tacked on or forced. Some trivial points may be treated in greater detail than important points, and there is usually some dead wood that might better be cut out.

Low: This paper starts anywhere and never gets anywhere. The main points are not clearly separated from one another, and they come in a random order--as though the student had not given any thought to what he intended to say before he sat down to write. The paper seems to start in one direction, then another, then another, until the reader is lost.

3. Wording

High. The writer uses a sprinkling of uncommon words or of familiar words in an uncommon setting. He shows an interest in words and in putting them together in slightly unusual ways. Some of his experiments with words may not quite come off, but this is such a promising trait in a young writer that a few mistakes may be forgiven. For the most part he uses words correctly, but he also uses them with imagination.

Middle. The writer is addicted to tired old phrases and hackneyed expressions. If you left a blank in one of his sentences, almost anyone could guess what word he would use at that point. He does not stop to think how to say something; he just says it in the same way as everyone else. A writer may also get a middle rating on this quality if he overdoes his experiments with uncommon words: if he always uses a big word when a little word would serve his purpose better.

Low. The writer uses words so carelessly or inexactly that he gets far too many wrong. These are not intentional experiments with words in which failure may be forgiven; they represent groping for words and using them without regard to their fitness. A paper written entirely in a childish vocabulary may also get a low rating, even if no word is clearly wrong.

4. Flavor

High. The writing sounds like a person, not a committee. The writer seems quite sincere and candid, and he writes about something he knows--often from personal experience. You could never mistake this writing for the writing of anyone else. Although the writer may play different roles in different papers, he does not put on airs. He is brave enough to reveal himself just as he is.

Middle. The writer usually tries to appear better or wiser than he really is. He tends to write lofty sentiments and broad generalities. He does not put in the little homely details that show that he knows what he is talking about. His writing tries to sound impressive. Sometimes it is impersonal and correct but colorless, without personal feeling or imagination.

Low. The writer reveals himself well enough but without meaning to. His thoughts and feelings are those of an uneducated person who does not realize how bad they sound. His way of expressing himself differs from standard

English, but it is not his personal style; it is the way uneducated people talk in the neighborhood in which he lives.

II. MECHANICS

5. Usage, Sentence Structure

High. There are no vulgar or "illiterate" errors in usage by present standards of informal written English, and there are very few errors in points that have been emphasized in class. The sentence structure is usually correct, even in varied and complicated sentence patterns.

Middle. There are a few serious errors in usage and several in points that have been emphasized in class, but not enough to obscure meaning. The sentence structure is usually correct in the more familiar sentence patterns, as in parallelism, subordination, consistency of tenses, reference of pronouns, etc.

Low. There are so many serious errors in usage and sentence structure that the paper is hard to understand.

6. Punctuation, Capitals, Abbreviations, Numbers

High. There are no serious violations of rules that have been taught--except slips of the pen. Note, however, that modern editors do not require commas after short introductory clauses, around nonrestrictive clauses, or between short coordinate clauses unless their emission leads to ambiguity or makes the sentence hard to read. Contractions are acceptable--often desirable.

Middle. There are several violations of rules that have been taught--as many as usually occur in the average paper.

Low. Basic punctuation is omitted or haphazard, resulting in fragments, run-on sentences, etc.

7. Spelling

High. Since this rating scale is most often used for test papers written in class when there is insufficient time to use the dictionary, spelling standards should be more lenient than for papers written at home. The high paper usually has not more than five misspellings, and

these occur in words that are hard to spell. The spelling is consistent: words are not spelled correctly in one sentence and misspelled in another, unless the misspelling appears to be a slip of the pen. If a poor paper has no misspellings, it gets a 5 in spelling.

Middle. There are several spelling errors in hard words and a few violations of basic spelling rules, but no more than one finds in the average paper.

Low. There are so many spelling errors that they interfere with comprehension.

8. Handwriting, Neatness

High. The handwriting is clear, attractive, and well spaced, and the rules of manuscript form have been observed.

Middle. The handwriting is average in legibility and attractiveness. There may be a few violations of rules for manuscript form if there is evidence of some care for the appearance of the page.

Low. The paper is sloppy in appearance and difficult to read.

APPENDIX B

PRETEST AND POST-TEST ASSIGNMENTS

APPENDIX B

PRETEST AND POST-TEST ASSIGNMENTS

"WANTED: AN INTERESTING EDUCATION"

Along with the many present problems that our country faces, there is a continuing problem that all but a few activists overlook. This is the need to make high school curriculum interesting. For unless we do this we will be a people characterized in the future for our narrowness, shallowness of thinking, lack of insight and perspective.

First of all, let us examine the deficiencies of present education. They are in the fields of relevance, time, rigidity, and interest.

In the field of relevance, many courses offered in present-day high schools follow the courses offered to our parents in the 30's. As such, these courses, both in variety and modern-day orientation, do not meet the needs and wants of present-day students.

Regarding time and rigidity, may I say that a typical high school schedule puts a strait-jacket upon a student who seeks to learn more. The typical high school schedule is the same, day after day, with no allowance for the student either to pursue individual interests or take extra courses or do independent study that he might want to do. All of these things are conducive to furthering one's education, but most present-day schedules hamstring the student from doing these things, thus inducing in him the feeling that he is doing only what the authorities will let him do and not what he may need. He, under this system, eventually loses all individual incentive and becomes an automaton. Thus rigidity produces boredom, resentment, and depression.

> But how about interest? This is obviously the most important part of the discussion, for if the individual student feels that he cannot make a meaningful contribution to the learning process, he will reject it. This is where dropouts and disillusionment come from. The traditional lecture format with the instructor telling the pupils that "such and such is true" without any justification is a disaster. Once again, it is 1929 methods in 1969 classrooms! The pupils are not part of the learning process, because they are being spoon fed everything and, told, in effect: "Don't ask too many questions."
>
> [Mark Gruenberg in How Old Will You Be in 1984?, ed. by Diane Divoky (New York: Avon Books, 1969), pp. 96-97.]

Mark Gruenberg was a high school student at the time that he wrote the above selection. He found his educational values in conflict with the program of his high school. In his argument, he states some general deficiencies of the educational program, as he perceived it.

Consider your education. What conflicts between your goals and the goals of your school have you experienced? Can you remember specific incidents which led to conflicts, either within yourself or in an outward confrontation with others?

Assignment. Write a narrative essay of personal experience, directed toward an audience of teenagers, identifying a specific incident which involved a conflict between you and the educational "establishment," and describe the factors involved in the conflict. Include an explanation of how the conflict was solved or why a solution was impossible.

MODE OF DISCOURSE: PERSONAL NARRATION (PRETEST)

"A PLEASANT CHILD. . .AN UNRULY TEENAGER"

> It is bewildering for a parent to watch a pleasant child turn into an unruly teenager. Suddenly, nothing suits him. The house is crummy, the car is junky, we are old-fashioned. His inner radar detects what irritates us most. If we value neatness, he will be sloppy. If we insist on good manners, he will interrupt conversations, use profanity and belch in company. If we enjoy language that has grace and nuance, he will speak slang. If we are concerned about health, he will smoke like a chimney and wear summer clothes in freezing weather. Buying a suit, he may ask the salesman, "If my parents like it, can I bring it back?"
>
> [Dr. Haim Ginott, Between Parent and Teenager, quoted in Invention, by John C. Adler (New York: Harcourt Brace Jovanovich, 1972), p. 140.]

In the above description, Dr. Ginott expresses some general attitudes about teenage behavior from the point of view of parents. Consider conflicts between you and your parents which centered on concerns about your behaviors.

Assignment. Write a narrative essay of personal experience directed toward an audience of teenagers. In the essay, present a specific incident involving a conflict between you and your parents. Describe the factors involved and in the conflict and explain how the conflict was solved or why a solution was impossible.

MODE OF DISCOURSE: PERSONAL NARRATION (POST-TEST)

"NONCONFORMITY"

Whoso would be a man, must be a nonconformist. He who would gather immortal palms must not be hindered by the name of goodness, but must explore if it be goodness. Nothing is at last sacred but the integrity of your own mind.

[From "Self-Reliance," by Ralph Waldo Emerson, as cited in Invention, by John C. Adler (New York: Harcourt Brace Jovanovich, Inc., 1972), p. 130.]

Remember always that you have not only the right to be an individual; you have an obligation to be one. You cannot make any useful contribution in life unless you do this.

[Eleanor Roosevelt as quoted in Great Words of Our Time, selected by Dee Danner Barwick (Kansas City, Missouri: Hallmark Cards, Inc., 1970), p. 13.]

My Beard

IS IT STRANGE IS IT FRIGHTENING TO SEE
DOES IT DERANGE OR BRING CHANGE
TO YOUR RULES OF CONFORMITY?

Why do you frown at my beard--
An item so dear to me?
Does it bring to mind
That I'm trying to find
A thing you may never see--
A thing from the past
To which I hold fast
Known as identity?

--Jimmie Sherman

[In Invention, by John C. Adler (New York: Harcourt Brace Jovanovich, Inc., 1972), p. 131.]

The three selections quoted above contain varying thoughts about individualism and nonconformity. Emerson says nonconformity is a requirement of manhood; Jimmie Sherman sees his beard as a symbol of his identity; Eleanor Roosevelt says we have an obligation to be an

individual. Think about what these three people are saying about nonconformity. Does Jimmie Sherman's beard make him a nonconformist? Does it help him to be an individual, a man?

Assignment. Write an expository essay for an audience of teenagers. In your essay, explain what nonconformity means to you. Weigh and describe its beneficial and detrimental aspects. Try to illustrate your major ideas with supporting examples.

MODE OF DISCOURSE: EXPOSITION (PRETEST)

"Fear"

For fools rush in where angels fear to tread.

[From "An Essay on Criticism," by Alexander Pope, quoted in The Oxford Dictionary of Quotations, Second Edition (London: Oxford University Press, 1955), p. 383, 5.]

. . . the only thing we have to fear is fear itself.

[Franklin Delano Roosevelt, First Inaugural Address, March 4, 1933, quoted in The Oxford Dictionary of Quotations, Second Edition (London: Oxford University Press, 1955), p. 408. 20.]

He has no hope who never had a fear.

[From Truth, by William Cowper, quoted in The Oxford Dictionary of Quotations, Second Edition (London: Oxford University Press, 1955), p. 164, 15.]

The above statements contain varying thoughts about fear. Pope gives the emotion of fear "angelic" qualities; Cowper says fear is a prerequisite of hope; yet, Roosevelt says fear should be eliminated. Think about what these three people are saying about fear. Do the three men mean the same or different things when they use the word fear?

Assignment. Write an expository essay for an audience of teenagers. In your essay, explain what fear means to you. Weigh and describe its beneficial and detrimental aspects. Try to illustrate your major ideas with supporting examples.

"WHO SHOULD BE HIRED?"

Once upon a time there was a land in which the Gleefs held control. All important posts were held by Gleefs. Schools were run by and for the Gleefs. Money, the chance to travel, the best foods, the most interesting uses of leisure time--all were reserved in large quantities for Gleefs. Being a Gleef was hereditary. It was impossible to become one.

The non-Gleefs, needless to say, lived totally different lives. They were called Yooms, and they did all the jobs rejected by the Gleefs. The ones who didn't work at all were put in jail or allowed to starve.

A change in government took the power from the Gleefs. The new governors found that the clerical jobs were still done most efficiently by the Gleefs. Captain Klojj is now hiring a staff for his office. Even though he has been told to make every effort to hire Yooms, as repayment for their previous years of suffering, he is tempted to hire Gleefs. They spell better, are more punctual, and more accustomed to working. The captain knows that his own promotion depends in part on his skill as a manager.

[Thelma Altshuler. Choices (Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970), p. 67.]

The above situation represents a basic problem which exists today. Form an opinion about what Captain Klojj should do and think about the ways in which the situation could be applied to today's society.

Assignment. Write an argumentative essay supporting one of the following points of view: "Employers should hire people on the basis of their ability to do the job," or "Employers should consider a man's past and his potential as well as the cheapest way to do the job."

[NOTE: An argumentative essay strongly supports one side of a controversial issue but also considers points in support of the opposing side of the issue. In order to

convince his audience to agree with him, the author of an argumentative essay supports his position with evidence and reasoning which outweighs that which may be presented in support of the opposing position. In your essay, it is not necessary that you refer to the situation presented in this assignment; but if you wish to refer to it, consider the fact that your readers will not be familiar with it and you will need to explain the situation.]

MODE OF DISCOURSE: ARGUMENTATION (PRETEST)

"WHO SHOULD PERFORM?"

Mr. Rivers is conducting tryouts for an extracurricular school play. Two girls are trying out for the female lead. Mary, a popular girl who is a leader in many school activities, reads the part extremely well. Mr. Rivers is certain that she could prepare for the role with very little coaching from him. Joan, a shy girl who has not been in the limelight in school activities, reads the part less well than Mary; Mr. Rivers predicts that she has the potential to play the role adequately after a great deal of coaching from him. Mr. Rivers must decide which girl to cast in the leading role. He realizes that the performance may be better with Mary in the role; on the other hand, he sees that casting Joan in the role could help her to increase her confidence in herself.

Mr. Rivers faces a situation which is common for many teachers who coach or advise extracurricular school activities. Form an opinion about what Mr. Rivers should do, and think about situations similar to this which might occur in other extracurricular activities.

Assignment. Write an argumentative essay supporting one of the following points of view: "The main purpose of an extracurricular activity is to provide participation in school activities for those students who are already skilled in the activity," or "Participation in school activities should be for those students who wish to learn or improve their skills in the activity."

[NOTE: An argumentative essay strongly supports one side of a controversial issue but also considers points in support of the opposing side of the issue. In order to convince his audience to agree with him, the author of an argumentative essay supports his position with evidence and reasoning which outweighs that which may be presented in support of the opposing position. In your essay, it is not necessary that you refer to the situation presented in

this assignment; but if you wish to refer to it, consider the fact that your readers will not be familiar with it and you will need to explain the situation.]

MODE OF DISCOURSE: ARGUMENTATION (POST-TEST)

APPENDIX C

VALIDATION AND APPLICATION

OF TEACHER RESPONSES

APPENDIX C

VALIDATION AND APPLICATION

OF TEACHER RESPONSES

DIRECTIONS - PLEASE READ CAREFULLY

Thirty-five sets of comments written in response to first drafts of student essays follow. These comments were selected randomly from all comments the investigator wrote in response to 224 first draft essays included in a study concerned with composition evaluation. The investigator attempted to use one of two kinds of evaluative comments in each set of comments. Please read the definitions which follow and then identify which type of comment each of the thirty-five sets of comments represents.

The directive response is a written comment which gives a specific direction to the student regarding the improvement of the writing skills which he is expected to demonstrate in his writing. Place a 1 on the blank next to each numbered response which is directive.

The clarifying response is a question or series of questions designed to help the student evaluate the nature of his ideas and consider alternatives in relationship to the writing skills he is expected to demonstrate in his writing. Place a 2 on the blank next to each numbered response which is clarifying.

1 1. You have chosen a topic which should be of interest to most students at Northeast this year. The assignment stated that you should identify a specific incident which involved a conflict. You have written an expository essay explaining the process which led to open campus for lunch. Choose one incident, perhaps a lunch period last year, and describe it in detail, explaining your feelings. Then, the open lunch period of this year can be presented as the resolution of the conflict.

2 2. Your essay reminded me of a similar situation I faced in college--a five-hour fine arts requirement. After struggling through two extremely frustrating art courses, I decided I couldn't tolerate another. I, however, had a more appealing alternative than you: a music appreciation travel course in New York City. Your essay communicates your dislike for physical education classes, but I'm not sure I know the specific aspects of these classes which cause you to dislike them. Could you describe a specific incident which includes most of the disagreeable aspects of physical education courses? What things would you emphasize in order to bring out these disagreeable aspects?

2 3. Your essay caused me to realize that something most teachers consider routine (in this case, assessing a fine for a lost book) may be upsetting for students. Your essay covers a long period of time. If you were to limit it to a shorter time period, what incident should be described in detail? If, for example, you chose to describe the time when you discovered that you had the wrong book, how might you include the necessary information which led up to the situation? Would a flashback technique be effective?

1 4. Your essay contains an explanation of several things which need to be changed in order to make education relevant for all students. You have given much consideration to these things. The assignment asked that you consider a specific incident which involved conflict. You stated that you could not remember a specific conflict; then, later in your essay, you give some clues that there were some incidents which involved conflicts, at least minor ones. For example, you stated that you were required to take five classes when you only needed three. Try to recreate that conversation with your counselor, revealing your feelings toward the requirement and toward the provision that allows work-study students to be exempted from that requirement. Consider how you can make your essay interesting to an audience of teenagers.

1 5. Your essay makes me more convinced than I was before reading it that school systems should stop labeling students according to ability. What a devastating effect this can have on a student! You describe two different incidents in which this practice of classifying students in terms of grades caused conflicts for you. Choose one of these incidents and expand upon it. Use descriptive details and dialogue to make the situation come alive. Reveal your feelings and reactions to the situation.

2 6. You have written a well-organized essay. Do you know someone who is "anti-everything?" Would using him as an example improve the third paragraph on page two? Can you use an example of a nonconformist who practices a "give-and-take" policy? Could you describe his actions? Why do some people regard nonconformists as trouble-makers? How can we judge that a person has gone too far with nonconformity? How many clichés can you find in your essay?

2 7. The final sentence of your essay expresses a very perceptive thought! Can you give some examples to illustrate the difference between a mere individual and a nonconformist? How can a person tell the difference between an immature person and a nonconformist? Can you think of examples which would make the second paragraph on the last page more vivid? How would a person cheat himself by saying he is a nonconformist when he is not?

1 8. You have pointed out some ways in which people can be nonconformists and still conform to the laws which are set up for the benefit of society. Read all of your sentences carefully, and make sure they express complete thoughts. In the last paragraph, give some examples of how nonconformity gives you more freedom. You will need to work hard on writing sentences which express complete thoughts. You may want to set up some time during a study hall for special help with this problem. You have some very good ideas which need to be expressed in better sentences.

1 9. Your idea that the nonconformist cannot be identified as belonging to one particular group is a good one. You take two pages to lead up to this idea. In these pages you make some general statements about demonstrators and the reactions of people to demonstrators. If you believe this information is necessary to illustrate the central point of your essay, try to condense it and to make the statements more concrete by using more examples. Some of your ideas are difficult to understand because of the way you use pronouns. At the bottom of page one, for example, you state: "I am sure this is probably not . . ." Explain what this refers to. Look for other similar uses of pronouns in your essay. Look at your word usage and make sure you have used the most appropriate, specific word in each case. For example, "like themselves" in the last paragraph might be replaced with "true individuals."

2 10. By taking the perspective of the manager in your essay, you were able to explain some of the basic problems which must be considered when hiring new employees. Does the "three-pronged" plan you propose call for the hiring of three different kinds of employees? If so, how could you explain this idea more clearly? Can you explain the method which opposes the three-pronged plan more fully? Won't the man with the potantial to do a job now have to be retrained in the future, just as the man with ability will have to be? Should you mention the story in the assignment without explaining it to your readers? Are you assuming that all of your readers have also read the assignment?

1 11. You have presented a very strong case in support of one side of the issue. In order to make your argument stronger, you should consider the fact that in the past members of minority groups who have had the same qualifications as others have been discriminated against. Women and other minority group members have been discouraged from aspiring to prepare for certain kinds of jobs. Include some advice for the employer who is faced with two applicants with equal ability to do a job, one of whom belongs to a minority group. It might also be helpful to state a case either for keeping the educational conditions for minority groups as they are, or for changing these conditions.

1 12. You have written a strong conclusion for your essay. Some of your paragraphs contain material which is not directly related to your topic. Your essay will be stronger if you eliminate this material. One example of this is the general information about maintaining society. Your essay would be more vivid if you were to use two hypothetical applicants for a job, one with ability and another with past experience, and contrast them as you develop the ideas in your essay.

1 13. You have chosen to relate an experience which had an impact upon your subsequent actions. Decide which part of the experience had the most impact upon you. Describe that part of it in detail, adding dialogue to make the scene more vivid. Describe your thoughts and feelings in detail. Try to use the flashback technique to include any background material which is essential.

1 14. You have described very well the feelings you had as you experienced the death of your cousin. Your experience covers quite a few days. Your essay would improve if you could center it on one event. If you were to set the beginning of the essay at the funeral, you could reveal all the necessary information about the events which led up to it through your thoughts. Try to combine your thoughts with the statements made during the funeral services. Try to make your dialogue more realistic.

2 15. In the conclusion of your essay, you have included a profound thought about human behavior. Can you think of a better way to lead into the statement of this thought, one which avoids the statement, "The main reason for remembering this . . .?" Why was it very important that you not be late for school? Could you describe your return trip to school in more detail, actually stating your thoughts as they occurred to you? Did you think that you would be late, or did you already know that you would arrive before the bell? Why should the paragraph which describes what you did after you returned from school be included? Could your essay just as well end with your arrival at school?

1 16. You have related very well the impact of your grandfather's death upon you. You have related your thoughts during the funeral very well. Try to describe the physical setting of the church. Put the speech of your sister and the minister in dialogue.

2 17. You have used some beautiful figurative language in this essay. Why did you decide not to identify the place specifically? In the last paragraph, what seems to be speaking, the wind or the grass? Who is the you in the last sentence? Would the addition of color add to the contrast between the bright and dull leaves you mention in the first paragraph?

2 18. Your topic is a good one. I've often wondered what a basketball game seen through the eyes of a player would be like. Your concluding paragraph includes a statement of the dominant impression of your essay--the monstrous image of Pershing Auditorium. Is everything you have included in your essay necessary in order to establish this impression? Is the bus trip and the description of the outside of the building necessary? What caused your jitters to disappear as the game progressed? Would describing some of the sounds in the auditorium add to the dominant impression you create?

1 19. Your ending is extremely appropriate for the particular impression you create in your description. You say that the pattern of your yard is interesting. Show us its pattern. In the paragraph about the heat on page two, describe the red-faced people in more detail. "The dust made the exact description of the truck hard to bring in site [sic]--describe what the dust did.

1 20. You have captured much of the impersonality of a public waiting place. You probably observed more about the young girl than her pink coat. Describe her facial features. Try to describe your "reason for coming" in more detail. Describe any thoughts you had about him while you were waiting.

2 21. A restaurant during rush hour is a very good subject for a descriptive essay. Could you center on one aspect of your work and describe that in detail? What do the customers look like? How do they behave? Is the noise level high? What does the combination of odors from the cooking food smell like? What is your duty? How do you feel about rush hour? Do you like it better than the slow time?

1 22. You have presented a good illustration of a serious cause of loneliness--the death of a person close to you. Try to make your loneliness at this time a more vivid illustration. Explain your relationship with your grandfather by presenting some examples of situations in which he helped you overcome feelings of loneliness. Explain the reasons why you felt others couldn't understand you in the ways he did.

1 23. You have presented a description of loneliness and also some good advice about how to cope with it. Look at your first sentence (loneliness is solitary). Change your definition so that the words on either side of is are the same part of speech. In your first paragraph, try to present an indication of what you intend to discuss in your entire essay.

2 24. Your idea that maturity develops gradually is a good idea. What aspects of maturity develop gradually? Could you group these ideas together and find examples to support each of them? Would this arrangement improve your essay? What kinds of examples could you use to expand your one-sentence paragraphs? What is "the maturity level that people imply?"

1 25. Your description of the kind of courage a mature person must have is very good. Try to give specific examples which help to distinguish this kind of courage from heroism. You have given some examples of crisis situations in which a person must become mature or break. Explain how a person who does not have these experiences can become mature. Use a specific examples to illustrate the ideas presented in the second full paragraph on page two. Try to add a concluding paragraph which sums up all the characteristics of maturity you present in your essay.

1 26. Your use of the two examples to make your point about the draft is good. It seems that Nick's and Bob's attitudes toward the draft are different. When you begin to summarize on page three, make sure you distinguish between the two views. You can do this and still draw the conclusion you do at the end.

2 27. The organizational structure of your essay is excellent! How can you make sentences like the following one more direct and forceful: "By making the dress at home a lot of money can be saved." When is passive voice appropriate? Can you find other sentences in which you use passive voice? "The total cost would probably be around ten to fifteen dollars." What does the word "around" mean? Have you used it appropriately here? Could you give detailed explanations of how a basic pattern could be changed into many unique garments?

2 28. Your introductory paragraph is excellent. Are the ideas in your concluding paragraph part of a summary or do they contribute to the development of a main point of your essay? Do they belong in a concluding paragraph?

1 29. Your topic should be of interest to most music enthusiasts, since you attempt to show relationships between classical and contemporary music. Try to make your thesis statement more concise by explaining the kinds of effects the two types of music have had upon each other. Make sure the main ideas in your paragraphs are developed. For example, explain how the communists have used classical music for propaganda. Explain how the movie 2001 shows that classical music is a part of the future.

1 30. I'm sure that the topic of your essay is of concern to many students at Northeast. In order to show the contrast between students' behaviors and their ideas, you might include some typical statements students have made about ecology. Explain how you have determined that the littering at Northeast is deliberate. Suggest some actions students could take to improve the ecology of the school.

1 31. Your essay shows that you have experienced some enjoyable times with cycles. Your essay seems to be about something different than your introductory sentence suggests. After reading your essay, I felt that you really wanted to express the idea that the media have created an unfair impression of motorcyclists. You might want to write an essay explaining why this impression is not justified.

1 32. You have chosen a topic of interest to many students and teachers. You need to seek out more information about t he operation of modular scheduling. Modular scheduling and open campus do not go together automatically. Modular scheduling does allow for classes to meet for varying lengths of time on different days. Perhaps some of your arguments should be centered upon the advantages and disadvantages of this kind of time scheduling. You may want to change your thesis statement to include both modular scheduling and open campus. Then your arguments for open campus could remain in your essay. Read your essay carefully and look for words left out of sentences.

1 33. You have identified a problem which probably concerns the majority of students at Northeast--smokers and nonsmokers. Your concern seems to be with the unpleasant atmosphere in the rest rooms. Alternatives to the establishment of a smoking lounge might be worthy of consideration. The legal problem created by the establishment of a smoking lounge seems insurmountable, as schools would be reluctant to defy the law. You might try to identify schools which have smoking lounges for students and discover how they deal with the legal problem. Your argument on page two that a smoking lounge would cut down on unexcused absences is unsupported. Try to verify it if you want to keep it in your essay. Your essay would be stronger if you were to present the opposition earlier in your essay.

2 34. Your topic is one which is of national concern. How would your law making marijuana legal be stated? How would this law remedy the varying penalties which exist from state to state? How would Mr. Koch's law replace legislation with decriminalization? Is your final statement logical? How many of your sentences can be revised to express ideas more clearly?

1 35. You have chosen a topic which is probably of interest to most viewers of commerical TV. I believe all cigarette advertising has been banned from TV. Make sure you relate all ideas in your essay to your thesis statement. Indicate which types of commercials are misleading. Explain how the five points in your rating system would be applied to indiate which commercials should be rejected.

Limitations of space made it impossible to reproduce the thirty-five essays for which the above responses were written; thus, two examples of initial drafts of student essays follow--one with directive responses, the other with clarifying responses. These are not intended as examples of excellent essays; rather, they are essays which, when combined with teacher responses, demonstrate effective use of directive and clarifying responses. The ratings of the revisions of both of these essays indicated substantial improvement in general writing performance.

Essay I: Directive Responses

It all started when I was in the second grade in a small one-room country school about this time of year. There were twenty kids in the whole school, each of the eight grades had from one to four kids in it. We had a recess or break from our studies, to go outside and play. And mostly, all the girls played together and us boys played football. As I said, I was a small second-grader and played against the giant of the eight-graders. In diving at one of the giants legs to knock him down, his knee cracked me in the right tempo of my head with an audible thump. It almost knocked me out of reality. As I laid lifeless-like face down on the hard ground, I faintly heard them asking me if I was alright. Their voices shaky in my failure to respond. They carried me gently to our teacher who called my mom. She came and got me and took me home.

There I laid on the couch to rest, to settle my scambled and aching head, the pounding in my head was painful.

Well I slept restlessly through that hazy afternoon, but when sleeping, I was free from the swisting and pounding in my head. I was up and around by 7 o'clock that night, but was restricted to slow motion movement for a quick move would almost blow my head off. I stayed home from school the next day, sleeping and watching TV. The second day I persuaded my mom to allow me to go to school.

Sharp blows to my tempos have given me, as the Dr. calls them, severe migrain headaches a few times since then. Some painful and some I just had to sleep off in a couple of hours or so. This is just a little background. The time I am about to describe is the one remember like a boy remembers his first bicycle.

I was in the seventh grade. My dad had bought me a go-cart. I rode it endlessly through the days of the summer on an oval track Dad built in the pasture of our farm. The pasture was as rough as a wash board and sometimes the go-cart didn't run right. But through experience, I could regulate the gas as not to flood the engine on the ruts.

Some friends of mine came over one Sunday afternoon and were riding the go-cart. At first, it ran alright, but it soon wouldn't run right. This, I knew, was taking all the fun out of it for them and was embarrassing me. So I jumped on a spot on the frame right beside the engine and over the back wheels where I could crouch down and control the gas while my friend drove. We were going around the rutted track at a good pace, enjoyable to my friend and proving it could run decent. We were going about 20 mph with me crouched on back.

But as we hit the sharp ruts, my foot slipped off the frame and in front of the back wheel. The wheel grabbed my foot, and as it pulled my leg and body over, my head slamed to the hardened ground. I hit unrestrained, sounding like water being thrown onto cement, as it happened quicker than my reflexes. I quickly got up, feeling very stupid at my act. I ran to get back on but knew I was again in the grip of my most painful enemy. My head was throbbing with unbearable degree. So I headed for the house, hopefully thinking I could sleep it off in 15 or 20 minutes and rejoin the group without then realizing my absence.

But as I stumbled to the house, I knew this was the worst. 75% of my vision was gone, destroyed by white spots that looked as if someone had just cut away the world and left empty space. My head was throbbing in unison with my heart to twice its size. My whole left side was going numb. I tryed to wake my arm and leg to no success.

I reach for the door handle, but had to shift my head to get it between the white spots. As I intered, Mom immediately knew what was wrong and took me to their bedroom. As I looked at her, a white spot hid her from me. I could see her black hair and her rounded chin, but the rest was cut away like paper, leaving empty space. I wanted to cry out, but it didn't hurt. It frustrated me, I wanted to break away but was too weak. It pinned me with an iron grip and beat my head to rub it in. I felt sick to my stomach, ready to throw-up but not needing to. Feeling like someone was standing on my stomach. And my tongue started to get numb along with my left side. It felt 3 times as big as before.

As mom talked to me, I swirmed in pain. Her words echoed through my head as loud as a shout and like inside a giant cane or tunnel. She said she would remember to talk softer, and put wet rags on my head. I had to sleep to escape my pain.

I slept death-like for 5 hrs., awakening to a quiet room. My mom was along the bed, her mouth straight, her eyes worried, as I moved. I tryed to sit up, but she quickly caught me and tenderly laid me back down. I powerlessly fell back to sleep.

I slept on and off for 2 more days, not caring about the rest of world.

On the fourth day, my head and eyes were cleared up and I went back to school feeling like a freed animal.

This is a good narrative essay, but it does not quite fit the description essay assignment. If you would like to keep this basic topic, you could center on the description of things you saw through your distorted vision after the go-cart accident. You could use a flashback technique to provide the necessary background information.

Capitalize upon the very good description of your distorted view of your mother, and add details concerned with your observation of the bedroom.

Essay II: Clarifying Responses

Nonconformity to me is the way a person describes his own interests and not that of his friends. The person feels he is an individual when he does not conform to society. Many people may not like what the nonconformist does or has. This is what makes him stick out in society.

Nonconformists are very hard to be sometimes because sometimes they are made fun of by people who disagree with what they agree with. The nonconformist can be ignored and talked about behind his back. This causes a very bad complication when people talk about you behind your back. Someone might say something about you that may not be true and you wouldn't even know about it. Sometime the nonconformist is unable to gain a good job because of his individuality. He sometimes can be disliked even before his personality is really known. This is just because of the way he appears. For example parents want to be proud of their kids so they make the boys get hair cuts when they don't want too. This causes the boy to loose his individuality that he needs.

There are also beneficial aspects to nonconformity. Nonconformity makes more individuals. This is why no one is exactly alike. It tends to push people away from being like others, which is what many people try to do. It also helps to make people more interesting.

Your decision to show the relationship between individuality and nonconformity was a good one. What do you mean by "describes his own interests . . .?" Can you think of examples of nonconformists who are objects of gossip? What are they like? What causes nonconformists to be rejected when they apply for jobs? Do all boys whose parents make them cut their hair lose their individuality?

APPENDIX D

MECHANICS PROBLEMS CHART

What are some possible alternatives to having long hair which would allow boys to express their individuality? How does nonconformity make people more interesting?

APPENDIX D

MECHANICS PROBLEMS CHART

Student _____ Hour _____

Type of Problem	Pretest 1	Pretest 2	Pretest 3	Personal Narration	Description	Definition	Exposition	Argumentation
1. Sentence Problems								
1.1 Fragments								
1.2 Run-On Sentences								
1.3 Irregular Tense Forms								
1.4 Agreement of Pronoun and Antecedent								
1.5 Agreement of Subject and Verb								
1.6 Pronoun Reference								
1.7 Dangling Modifiers								
1.8 Misplaced Modifiers								
1.9 Overuse of Expletive								
1.10 Misuse of Passive Voice								

Student _____ Hour _____

Type of Problem	Pretest 1	Pretest 2	Pretest 3	Personal Narration	Description	Definition	Exposition	Argumentation
1.11 Sequence of Tense								
1.12 Sentence Variety								
1.13 Parallelism								
1.14 Other Sentence Problems:								
2. Diction Problems								
2.1 Cliches								
2.2 Wrong Words								
2.3 Abstract Words								
2.4 Pompous Words								
2.5 Deadwood								
2.6 Impoverished Vocabulary								
2.7 Overuse of Adjectives and Adverbs								
2.8 Other Diction Problems:								
3. Capitalization Problems:								
4. Spelling Problems:								
5. Punctuation Problems:								
5.1 Periods								

Student _____ Hour _____

Type of Problem	Pretest 1	Pretest 2	Pretest 3	Personal Narration	Description	Definition	Exposition	Argumentation
5.2 Question Marks								
5.3 Exclamation Marks								
5.4 Commas								
5.5 Apostrophes								
5.6 Hyphens								
5.7 Italics								
5.8 Quotation Marks								
5.9 Dashes								
5.10 Parentheses								
5.11 Brackets								
5.12 Semicolons								
5.13 Colons								
6. Other Mechanical Problems (Specify):								
6.1 _____								
6.2 _____								
6.3 _____								
6.4 _____								

RECORD OF STUDENT WORK ON MECHANICAL PROBLEMS

Student _____ Hour _____

Code Number	Name of Problem	Date Begun	Date Completed

APPENDIX E

COMPOSITION I

COURSE OUTLINE

APPENDIX E

COMPOSITION I

COURSE OUTLINE

Lincoln Northeast High School

1972-1973

Course Objectives

The development of writing skills usually leads to many desirable academic outcomes. Students who are skillful in writing usually do better in most academic subjects than students who lack writing skills. Yet, another perhaps more important benefit comes with increased skills in writing. Through writing, people can examine their experiences, give them form and shape, and thus increase their understanding of themselves and of life. This is the major objective of the course. Hopefully, as you work toward this objective, you will increase your abilities to think critically; to clarify, organize, and support your ideas; and to express your ideas in a mechanical written form which helps others to understand them clearly.

Learning Activities

During the first few days of the semester, you will write in class three different kinds of essays. These essays will give your teacher information about your present skills in writing and aid her in discovering individual learning activities which should help you to increase your writing skills. After comments have been written on these essays you will revise them once in class. After this first revision, you may, at your own option, revise them further.

After evaluating these essays, you and your teacher will plan an individual program for attacking your mechanical writing problems. These mechanics problems assignments

will help you to master one problem at a time throughout the course of the semester. The number of these mechanics problems you work on and the amount of effort you put into these assignments will be totally up to you. Increasing your mechanical writing skills will help you to communicate your ideas more clearly, but it will not transform superficial ideas into meaningful ones. Thus, giving form and shape to your ideas should always be more important than achieving mechanical perfection in your writing.

The course will be centered on five major units of study: Personal Narration, Description, Definition, Exposition, and Argumentation. The units will contain learning activities which will help you to write these five different kinds of composition. The most important assignments, then, are the five compositions which will be written in conjunction with these units.

Learning activities for each of the units will include assigned readings, large and small group discussions, and short written assignments. One of the assumptions of the teacher is that people write for an audience. The audience for your compositions will be your classmates as well as your teacher. After you have written and revised your essays, you will be expected to present them to a small group of your classmates and to participate in the evaluation of the essays of your classmates. Throughout the course of the semester, these experiences in evaluation should help you to learn how to provide meaningful feedback for others and to consider how your audience may respond to your writing.

Near the end of the semester you will again write three essays on topics similar to those assigned for the three essays at the beginning of the semester. You will not revise these essays. They will be used as a measure of your learning in the class.

Evaluation Procedures

The teacher will write on each of your essays to be revised comments which will help you in the revision process. No mechanical errors (errors in spelling, punctuation, usage, etc.) will be indicated on your papers. Instead, the teacher will periodically record these errors on a special form which will allow assessment of your progress in mechanical writing skills. No grades will be recorded on the essays. What your teacher will provide, instead of grades, is assistance designed to help you revise your essays as many times as you wish until you are satisfied with them.

All of your essays will be evaluated according to the Diederich scale which is included in this handout. On some occasions, your essays will be rated by other experts in the teaching of written composition; when this occurs, they will also be using the Diederich scale for their evaluations. Please study the evaluation scale and evaluate your compositions according to it. Ask other students to use it to rate some of your essays.

Your teacher will keep a record of the rating scores for each of the eleven initial compositions and for each of the eight revisions. Your nine-week evaluation will be recorded either "satisfactory" or "unsatisfactory." All students who have completed the assigned work will receive a "satisfactory"; all others, an "unsatisfactory." In addition, your teacher will prepare for you and for your parents a written evaluation of your progress in the course which identifies your major strengths and weaknesses in written composition as demonstrated by your work during the first nine weeks.

At the end of the semester, the teacher will evaluate each student's work in a written report which will identify his strengths and weaknesses. This report will also indicate the change, if any, which has occurred in his writing skills. To receive a passing grade, a student must have completed all nineteen composition assignments. The grades of students who have completed all nineteen assignments will be determined in one of two ways. Each student's performance on the three final assignments will be compared to his work on the three initial assignments. The difference in scores will be compared to the average difference which occurred for other students enrolled in Composition I. Students will then be classified into six groups which will be determined on the basis of the "growth" shown. These six groups will be equivalent to the six levels of passing performance provided for by the grading system of the Lincoln Public Schools. In addition to these measurements, each student's work as demonstrated on the three final assignments will be compared to the work of his fellow students; again, six classifications will be formed. Thus, for each student, there will be a measure of "growth" when his end performance is compared to his beginning performance; and, for each student, there will be a measure of his performance on the final assignments as compared to the performance of other students enrolled in composition. A combination of scores for these two categories will be used in the assignment of a student's semester grade.

Students at Northeast High School have the opportunity to elect pass-fail evaluation for a proportion of their course hours. Your teacher strongly recommends this method of evaluation for Composition I.

Hopefully, the evaluation process used in this course will be a valuable aid to your learning and will help you to develop skills in self-evaluation.

Preparation of Compositions

The following instructions apply to the preparation of the nineteen major writing assignments:

1. Write only on paper which the teacher provides for you.
2. Use a black ink pen (ball point or fountain). If you typically write with light pressure applied to the paper, you should probably use a medium point pen.
3. Include the following information on a cover sheet for your essay:

    a. your name
    b. your student number
    c. Composition I
    d. Mrs. Kelley
    e. the description of the assignment
    f. Hour (1, 2, 3, or 5)

4. Place only the page number in the upper right hand corner of each page of your essay.
5. Make sure that you leave a one-inch margin on the left side of each page. (You may wish to prepare a sheet with a heavy black line to use as a guide sheet.) Leave a margin of at least one-half inch on the right side. Write on only one side of the paper.
6. When you are finished writing your essay, arrange the pages in order with the cover sheet on top and fasten them together with a paper clip. (Do not staple them.)

Conclusion

Your teacher has attempted to plan meaningful learning experiences for Composition I. All assignments are intended to contribute to the improvement of your writing skills; none are intended as busy work. The improvement of your writing skills will depend, to a great

extent, upon the amount of effort you choose to devote to the assignments. Please feel free to call upon your teacher for individual help at any time. Also, please feel free to offer feedback about the class at any time. Your evaluation will help her to improve the course.