

RESEARCH STUDIES IN WRITTEN COMPOSITION

Lois V. Arnold
Pinellas County

Recently, while conducting a study of the effects of frequency of writing and intensity of teacher evaluation upon high school students' performance in written composition (2), the author searched the literature for carefully designed studies in the area of written composition. Professional journals contained many articles describing various techniques and procedures for instruction in writing and summarizing loosely controlled studies in written composition. However, in few of the studies is performance measured and are results analyzed objectively.

It is true that researchers encounter many problems in conducting studies of student performance in writing. Ingrid Strom (21) summarizes these difficulties as follows: "In evaluating the outcomes of instruction in the English language arts, the investigator has been faced with the problem of controlling variations in interest, sex, intelligence, chronological age, mental age, academic standing, race or nationality, socio-economic status, administrative devices for grouping students, and the climate of the school and the community. In an attempt to evaluate behavioral changes brought about in students as a result of instruction in English, researchers have carried out numerous normative surveys but few significant developmental and experimental studies."

Even with these existing problems, certain carefully designed research was located which proved relevant to the author's study of the effects of writing practice and teacher evaluation upon students' performance in written composition.

It is at the college level that there have been more experimental studies of the results of frequent writing practice by students. At Purdue University, Lokke and Wykoff (16) studied performance in freshman composition. During the semester students in two small classes wrote from 32 to 34 themes, with matching students in two control groups writing only half as many. In the experimental classes students devoted more class time to actual writing and discussion of their work. The authors conclude that the groups which did the greatest

amount of writing had fewer failures and gave evidence of greater improvement. The researchers note that approximately 40% of the students apparently reached their limit of achievement after writing twelve to fifteen themes.

At Michigan State College, Dressel, Schmid, and Kincaid (8) made a comparative study of those freshmen who had the most and those who had the least essay-type writing. Using a questionnaire-type method to ascertain the amount of time spent in writing, the researchers employed as criterion tests 300-word themes written at the beginning and at the end of the semester. It is the conclusion of the authors that many college freshmen do little written composition except in the language communication courses and that mere frequency of writing is not enough to improve the quality of writing.

Maize (17) reports a more recent experiment at Purdue which involved students in the lower fourth of the freshman class. These freshmen, with an average I. Q. of 99, were reading, on the average, at the tenth-grade level. Students in the control group used a grammar workbook-drill method and wrote weekly themes of approximately 250 words. Students in the experimental group wrote daily themes in a laboratory situation. No papers for this group were graded outside the class. Divided into groups of five, students read each other's work, with the teacher spending some time with each group. In addition, the teacher used the opaque projector and mimeographed the best papers. Only one objective was stressed at a time, such as the topic sentence, organization, or the conclusion. With the experimental group there was overwhelming evidence of superiority over the control group in the use of language.

When an emergency course was organized at the University of Alberta for the immediate training of teachers for Alberta schools, Buxton (5) was enabled to study effects of writing frequency and guided practice upon students' skill. His three treatments included 257 freshmen. A control group followed the regular program and received normal assignments in writing. Students in a "writing" group followed the same program of studies, but they wrote an additional essay of approximately 500 words each week. Students in the "writing and revision" group followed the same program as the "writing" group, but their papers received intensive marking and grading and students discussed at length corrections and revisions. Although there are some significant differences when individual categories on criterion tests are considered, the statistical analysis indicates that the performance of the three groups did not differ significantly.

More recently there have been studies reported concerning the effect of added writing practice upon the performance of high school students. In his study of those high schools from which superior students have come, Hook (13) notes that such schools require considerable writing on the part of their students.

Sauer (20) reports favorably the lay reader plan, in which housewives or other persons in the community with degrees in English or professional training in written composition help with the correcting of students' papers. He describes the contract correcting project involving the high schools in Newton and Concord and in the junior high schools of Quincy, Massachusetts, 1957-61. He concludes that after teachers and readers have cooperatively worked out certain problems, the lay reader does relieve the teacher. Generally, teachers, readers, and students indicated approval of the contract reader plan used in the schools Sauer describes.

In studying the effects of additional theme-writing, Peterson (19) considered achievement of senior high school students in areas of punctuation, paragraph identification, spelling, capitalization, and sentence structure. Through the use of lay-theme readers he also considered actual theme writing. He finds no significant differences between his control and his experimental groups. Peterson recommends a similar study covering a longer period of time and including not only themes, but all written work.

Heys (10) reports a study at Lincoln-Sudbury Regional High School with classes matched at grade levels nine through twelve. At each level one class wrote the equivalent of a theme a week, which was rigorously corrected, revised, and rewritten. Another class wrote themes every third week but spent one class period each week doing free reading. Criterion tests were two test compositions read by experienced readers and STEP Writing Tests. Heys generalizes that frequent writing practice probably results in greater improvement in grade twelve and with low groups. With the low groups, this improvement seems more apparent in the area of content and organization than in the area of mechanics or of diction and rhetoric. It is the conclusion of participating teachers that the experiment does not substantiate the claim that frequent writing practice results in improved written expression. In addition they conclude reading exerts a positive influence on writing ability, a factor separate from teacher personality and enthusiasm.

The theory that frequent writing practice improves writing is still a very positive factor in determining instructional procedures, even

though studies such as the one by Heys are beginning to question this assumption. French (9) found that in the period from 1947 to 1957 there had been a slight increase in writing, with students at the twelfth grade level and students who were college preparatory or advanced receiving the greatest amount of writing practice.

An important part of instruction in writing is of course evaluation, a term which carries a variety of meanings. Some teachers may regard evaluation as indicating errors in mechanics, commenting on the general effectiveness of a theme, and assigning a specific grade of A, B, C, D, or F. Other teachers may consider evaluation a process of conferring with the student and helping him note his own progress by comparing papers kept in his file folder.

The eternal question in evaluation is whether the hours spent in reading students' papers are producing the expected results. Loban, Ryan, and Squire (15) report that a study by Augustine Confrey of compositions corrected by teachers of college freshmen "showed that only twenty percent of the comments helped the writers become more self-directive, and almost thirty-six percent of the comments were worthless or positively false."

An additional problem in measurement of written composition is that of objectivity. To be valid, a test of student performance in written expression should require that the student express himself in writing. However, essay tests introduce such variables as rater subjectivity and interrelationships of the writer's experiences and attitudes toward composition topics. Attempts to control these variables have resulted primarily in research concerning the development and use of evaluation scales and the reliability of rating.

The Hillegas Composition Scale (11), which assigned values to various sample compositions, represents one of the early attempts (1912) to devise a means for measuring the general merit of composition. Although this study did not attempt to define merit, the sample compositions provided did give guidance on a comparative basis to teachers in making judgments concerning form and general effectiveness.

An experiment with the Hillegas Scale, which showed that the use of an objective measure such as this helped to unify grades given by various teachers to compositions, resulted in Ballou's Harvard-Newton Scales (4) in 1914. This instrument contained separate scales for measurement of four forms of discourse, with each scale containing six compositions. However, Hinton (12) reports there were insufficient samples of low quality. He also reports that a year later, by

adding several samples at different points at the middle part of the scale and inserting longer samples in place of the shorter, Thorndike developed the Extension of the Hillegas Scale. Hinton claims that these additions posed additional problems for the inexperienced scorer.

During the years intervening between the Hillegas Scale and the present time, there have been various other scales devised. Asbaugh (3) feels that the Willing Scale added much to composition measurement. This particular scale was among the first to provide for the separation of the element of thought from the mechanics of capitalization, punctuation, spelling, and paragraphing. Hinton (12) lists other scales, such as the Van Wagenen and the Hosic Scales, which included as a part of measurement not only mechanics, but structure, style, and thought content.

A recent study by the California Association of Teachers of English (6) resulted in the development of a six-point scale, based on levels of competence demonstrated in papers written by high school seniors in California. In this study of writing ability of the high school seniors, ten experienced readers at the University of California read and screened 561 essays.

The majority of the studies cited indicated that under certain conditions the use of composition scales tends to make more uniform the evaluation of written expression. Conditions specified include careful training of raters, sufficient composition samples, carefully devised comments and corrections, and pupil access to and familiarity with the scales.

Even though evaluation scales with sample essays and correction guides are utilized, there are two problems basic to the measurement of student performance on an essay-type test. One is concerned with reliability of student performance and the other with reliability of readers.

Diederich (7) suggests that an essay test is valid if students are required to do only that writing which is similar to that normally done and if marking takes into consideration not only content but additional factors. He also maintains that for a reliable estimate of a student's ability, there should be at least two sample compositions.

Diederich and other authorities have found that essays marked by as many as ten different teachers may be found to have grades ranging the full scale from A to F. Such disagreement among raters has resulted in various studies of reader reliability. Anderson and Thaxler

(1) found in their studies indications that a single reader who is trained can read essay tests with high reliability, and several readers will, with training, assign to essay tests scores which do not vary greatly. In a study of ratings given freshman compositions, Diederich, French, and Carlton (14) conclude that readers were not in agreement and that actually little is known about how to test by objective tests or by essay tests for such qualities as ideas, structures, style, and originality.

The essay part of the College Board Examination has evoked much discussion. Palmer (18) reports experimental administration of the General Composition Test in 1951, 1952, and 1953. During the next three years, the two-hour test was administered as a regular examination. According to Palmer the project was a failure because of low reader reliability, unsatisfactory validity, and impracticality of grading. Matching scores against students' high school course grades and teacher estimates of students' writing ability, he found the General Composition Test less reliable than the English Composition Test, which is largely objective, and the verbal section of the Scholastic Aptitude Test, which is wholly objective.

Studies of reliability of the reader and of pupil performance seem to reach agreement on one assumption. Anderson and Thaxler (1) summarize this agreement by pointing out that on the average, "... pupils do improve in their ability to use written English as they advance through secondary school."

Both in the review of the literature and in her own investigation, the researcher has been primarily concerned with writing frequency and teacher evaluation and their effects upon high school students' performance in written composition. Most of the studies cited fail to substantiate the claim that writing practice automatically results in improvement. Present evaluation procedures are also placed in question, with some credit accorded rating scales for their tendency to make grading practices more uniform.

The many questions concerning the effects of writing practice and evaluation influenced the researcher to study in particular these factors assumed to be so essential in the teaching of composition. Designed to include eight tenth grade classes with two teachers following four different approaches to intensity of teacher evaluation and frequency of writing, the author's study advanced and tested as main hypotheses that there are no significant differences resulting from writing practice or from variation in evaluation.

The four approaches included the following:

1. Students in the first group wrote infrequently, three 250-500 word themes each semester, and teachers practiced moderate evaluation.
2. Students in the second group wrote frequently, a sentence or paragraph each day, and teachers were moderate in their evaluations.
3. Students in the third group wrote infrequently, but teachers evaluated their themes intensively, marking all errors and writing extensive comments. These students were required to correct, revise, and rewrite.
4. Students in the fourth group wrote frequently, at least one 250 word theme each week, and teachers evaluated their papers intensively, requiring correction, revision, and rewriting.

Initially classified according to ability on the basis of verbal aptitude scores on the Differential Aptitude Test, students in each of the four groups were given alternate forms of STEP Essay and Writing Tests in the fall of 1961 and in the spring of 1962. Three certified and experienced teachers of English served as essay test raters.

A 2x2x3 design was used for analysis, with moderate and intensive evaluation, infrequent and frequent writing, and levels of ability analyzed as main effects. Two covariates, teacher-school and sex, were included as first and second orders of interaction. Results of tests were subjected to analysis of covariance. No significant main effects or interactions are shown by F tests.

The researcher finds the results of this study in agreement with the majority of research studies reviewed and concludes there is no substantiation for the claims:

1. That intensive and detailed evaluation by the teacher is more effective than moderate evaluation, which might stress one specific skill or concept at a time.
2. That frequent practice in itself effects writing improvement.

3. That any one combination of writing frequency and teacher evaluation is more effective than another.
4. That any combination of practice and evaluation is more effective for one ability level than another.

Findings of this study and of research reviewed accentuate the need for carefully designed research in written composition. With various studies discrediting the well-established assumptions that improvement in written composition results if students receive frequent practice and if themes are given careful and detailed correction, additional research is needed. Much more information is needed as to what happens to students of all ranges of ability, not only after themes have been written, but before and while students write.

Bibliography

1. Anderson, Harold A., and Thaxler, Arthur E. Reliability of reading an English essay test: a second study. School Review, 48 (September, 1940), 521-530.
2. Arnold, Lois V. Effects of frequency of writing and intensity of teacher evaluation upon performance in written composition of tenth grade students. Unpublished doctoral dissertation, Florida State University, 1963.
3. Ashbaugh, E. J. The measurement of language: what is measured and its significance. Journal of Educational Research, 4 (June, 1921), 32-39.
4. Ballou, Frank W. Scales for the Measurement of English Compositions. Cambridge, Massachusetts: Harvard University, 1914.
5. Buxton, Earl William. An experiment to test the effects of writing frequency and guided practice upon students' skill in written expression. Unpublished doctoral dissertation, Stanford University, 1958.
6. California Association of Teachers of English. A Scale for Evaluation of High School Student Essays. Champaign, Illinois: National Council of Teachers of English, 1960.
7. Diederich, Paul B. Measurement of skill in writing. School Review, 54 (December, 1946), 584-592.
8. 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, 46 (December, 1952), 285-293.
9. French, John W. What English teachers think of essay testing. The English Journal, 46 (April, 1957), 196-201.
10. Heys, Frank, Jr. The theme-a-week assumption: a report of an experiment. The English Journal, 51 (May, 1962), 320-322.
11. Hillegas, Milo B. A Scale for the Measurement of Quality in English Composition by Young People. New York: Teachers College, Columbia University, 1913.

12. Hinton, Eugene Mark. An analytical study of the qualities of style and rhetoric found in English compositions. Unpublished doctoral dissertation, Teachers College, Columbia University, 1940.
13. Hook, J. N. Characteristics of award-winning high schools. The English Journal, 50 (January, 1961), 9-15.
14. Judges disagree on qualities that characterize good writing. High Points, 44 (February, 1962), 75-76.
15. Loban, Walter, Ryan, Margaret, and Squire, James R. Teaching Language and Literature. New York: Harcourt, Brace & World, Inc., 1961.
16. Lokke, Virginia, and Wykoff, George. Doubling writing in freshman composition experiment. School and Society, 48 (November, 1948), 437-439.
17. Maize, Ray C. Theme a day. National Education Association Journal, 42 (September, 1953), 335-336.
18. Palmer, Orville. Sense or nonsense? The objective testing of English composition. The English Journal, 50 (May, 1961), 314-320.
19. Peterson, E. Raymond. Achievement effect of theme and paragraph writing in senior high school through use of lay-readers. Unpublished doctoral dissertation, University of Kansas, 1961.
20. Sauer, Edwin H. Contract Correcting: The Use of Lay Readers in the High School Composition Program. Report to the Cooperative Research Program of the United States Department of Health, Education, and Welfare, and to the School and University Program for Research and Development. Cambridge, Massachusetts: The Harvard University Press, 1962.
21. Strom, Ingrid M. Summary of investigations relating to the English language arts in secondary education: 1960-1961. The English Journal, 51 (February, 1962), 123-140.