CRITERIA FOR GRADING STUDENT WRITING QUALITY:

A SYNTHESIS OF THREE EMPIRICAL APPROACHES

TO ASSESSING WRITING QUALITY

Richard F. Thompson

Northern Virginia Community College

The foundation to instructional planning in composition courses has not

been laid. The literature does not provide a coherent, agreed-upon set of

criteria that defines good writing. It is not that teachers cannot recog-

nize quality; for years researchers (Diederich, 1967, 1974; Godshalk et al,

1966) have shown that English teachers can consistently agree in their

rapid, holistic sorting of papers of differing quality on the basis of a

general impression. What we do not know is the decisive qualities on which

the sorting was based. Efforts to identify these qualities have taken two

major thrusts: a priori--such as constructing rating scales, course objec-

tives, or grading criteria reasoned from scholarly or intellectual discussion,

and empirical analyses--studies relating computer-detectable, syntactical,

content, or mechanical attributes to holistic, teacher evaluations of student

essays. Differences in terminologies and in conclusions make the identifi-

cation of student writing qualities even more complex. How does one recon-

cile the computer analysts' (Slotnik, 1971; Page and Paulus, 1968) variables

of different words, paragraph length, or word length to the syntax analysts'

(Combs, 1976, 1977; Hunt, 1979; O'Hare, 1973; and Mellon, 1969) variables of words per clause or independent clause, or to the content and mechanics analysts' (Butler, 1936; Cohen, 1971; Diederich, 1974; Slotnik, 1972; and Thompson, 1976) variables of supportive detail, cogent reasoning, organization, and mechanics?

What is needed is a synthesis of empirical approaches to provide clarity and a basis for teacher discussion of essay grading. In this study, variables from computerized grading of essays, as well as from sentence combining, mechanical and content analyses have been examined and compared for the first time in the same study to the writing quality of students' papers. Statistical comparisons identify four variables which can dependably and reasonably explain the teachers' grading of these essays. Results are examined in light of traditional criteria of good writing. On the basis of this examination, a recommended checklist of variables is proposed to help teachers in their consideration of student writing.

Literature Review

The need for this study stems primarily from problems and inconsistencies in rating scales of, and essays on, writing quality. Teachers like Roberts (1977) or writers like Orwell (1977) have emphasized what they in their experiences have thought to constitute writing quality, and these essays have found their way into readers for composition students. On the one

78

hand, Roberts advises students to avoid obvious content, abstractions, excessive padding, pat expressions, colored and colorless words. On the other hand, Orwell recommends that one avoid pretentious diction, filler words or phrases, meaningless words, and dying metaphors. In addition, he suggests that one write from mental pictures. But these authors, writing near the end of distinguished careers, intended more to provide helpful advice to improving writing rather than to provide a coherent statement of writing quality. Since they did not refer to each others' advice, they left to others the task of sifting through their essays to discover agreements, gaps, and differences.

To some degree, researchers have begun this sifting as they have constructed writing scales to aid teachers or readers in evaluating student themes. Most scales [like Cohen's (1971), the Cleveland Composition Rating Scale (Follman and Anderson, 1967), or the California Essay Scale (Follman and Anderson, 1967)] list fourteen to twenty writing attributes. This list might include such guidelines as Clarity, Structural Variety, Logical Effectiveness, Unity, Organization, Mechanics, Content, Style, Ideas, Specificity, and Rhetorical Force. But the scales can confuse or confound one writing criterion with another. For example, in the California Essay Scale the criterion ideas and their development under Organization seems to repeat the question of supportive evidence and completely presented ideas under Content. The California scale, and others, could probably be shortened and thus made easier to use.

Further, researchers have not established that the attributes listed in the scales either predict writing quality or are used in a teacher's evaluation. In one study, Klein and Grover (1970) factor analyzed a rating scale with ten criteria. They found that the readers' scores on each rating reduced to one factor--that a measure of one rating was a rating of another. This result raised the question of whether writing quality is comprised of only one attribute or whether teachers grade each section of a writing scale on the basis of a holistic rating. Rating scales, then, may not be very useful for determining the teachers' bases for grading papers. A more direct approach is needed.

The National Assessment of Educational Progress (Slotnik, 1972) provided further evidence. In one study, mechanical errors were detailed and counted in students' papers that had been graded previously and holistically. Writing experts compared the levels of writing quality to the incidences of mechanical errors such as comma errors, run-on sentences, subject/verb disagreement, and usage errors. On the basis of this comparison, the experts characterized the mechanical errors made at each level of writing quality. Among other things, they found that writing quality was inversely related to the incidences of mechanical errors: the better the essay, the fewer the mechanical errors.

More recent research has shown an empirical relationship of some variables to writing quality. Thompson (1976) found that errors in Unity, Unsupported

80

Statements, Cogency, Coherence, and Wordiness validly predicted teachers'

assessment of writing quality in independent sets of papers. Slotnik

(1971) as well as Page and Paulus (1968) have shown that computer variables

also predict writing quality. O'Hare (1973) and Combs (1976, 1977) have

shown that subordination seems to influence teachers' assessment of writing

quality. The issue to be addressed next is whether these three approaches

can be synthesized to give an adequate, comprehensive set of attributes

that can predict teachers' assessment of writing quality.

Method

To make such a synthesis, variables from these three analytical approaches

were identified in student papers and related to writing quality in a

stepwise regression procedure. Thirty-six papers were drawn randomly

from the last papers students wrote in the Winter, Spring, and Summer

Quarters at Northern Virginia Community College. Student papers came

from classes with the same teacher, same teaching and grading strategy,

and the same writing assignments. Moreover, these students represented

a cross section of the student population at Northern Virginia Community

College. This population consists of both men and women of widely varying

ages and background experiences. The average age is 27, and more than

half of the students are employed either full or part time.

Students wrote, as assigned, a critical review of Peter Drucker's (1973)

"The Surprising Seventies," agreeing or disagreeing with Drucker's ideas

and providing some support for their comments. Three readers then holistically rated these papers on a scale of one to four, with four being the highest mark. The inter-reader reliability was .86, and the papers ranged in quality from very poor to very good.

The papers were then submitted to three kinds of analyses. In the first analysis two readers examined the paper for Content and Mechanical errors. Both readers marked student errors on a copy; readers then compared analyses and discussed differences. If the two readers disagreed over the existence of an error or disagreed in its classification, the error was not counted. After these discussions, student errors in Unity, Support, Cogency, Coherence, Conciseness, and Mechanics were counted and listed under each of these categories.

These error counts had to be weighted since students wrote papers of different lengths. In this study errors were standardized by the number of paragraph units a student wrote. This standardization took two steps. First, the word length of each theme was divided by 75 (the mean paragraph length of all themes). The result was used as the number of paragraph units a student wrote. Second, the number of errors in each category was divided by the number of paragraph units. The result is the ratio of errors to paragraph units; this ratio was roughly equivalent to the number of errors per paragraph. For example, in Table 1, the student wrote 456 words and thus 6.09 paragraph units (456/75=6.09). Each error thus was divided by 6.09 (1/6.09=.16, 3/6.09=.49...10/6.09=1.64).

TABLE 1: SAMPLE OF ONE STUDENT'S CONTENT AND MECHANICS PROFILE

Type of Error	Unity	Support	Cogency	Coher	Concis	Mechan
Number of Errors	1	3	1	0	0	10

Number of Words = 456   Number of 75 word paragraph units (456/75 = 6.09)

| Ratio Errors/ 75 paragraph units | .16 | .49 | .16 | 0 | 0 | 1.64 |

The second analysis of the papers--the computer analysis--replicated, in part, Slotnik's (1971) study. A program, in BASIC, was written to read and count the objective variables. A stepwise regression analysis showed nine variables as significantly predicting the holistic grade: the number of words, different words, sentences, and paragraphs; the paragraph's length in sentences and words; the standard deviation of the sentences' length; the average length of the words in letters; and the ratio of words to different words. These computer variables were later compared to the content and mechanics variables as well as to the variables from the third analysis--sentence-combining.

The sentence-combining variables were found with one reader (and afterwards checked by two other readers) counting all the T-units (the independent clauses and all its dependent phrases and clauses) as well as the number of clauses. The number of T-units and the number of clauses were divided by the number of words in the essay. T-units and clauses were examined since Combs (1976, 1977) has set a precedent following Hunt's (1970) recommendation that these two variables best indicate sentence maturity.

83

one additional variable was added: an index of clause subordination. This index, the ratio of clauses to T-units, was used to reflect students' heavy reliance upon the clause as the means of subordinating an idea.

Thus, student papers had six Content and Mechanical variables, nine computer variables, and three sentence-combining variables of subordination. All of these variables were examined in one stepwise-regression analysis to determine which significantly contributed to a set of variables that could dependably predict the holistic score of the papers, the quality of the papers.

Data Analysis

Table 2 lists the results of the regression analysis.

The findings of this study to some degree reinforce the findings of an earlier study (Thompson, 1976) in which Unsupported Statements, Unity, and Cogency errors dependably predict reader grading. Although all six variables in Table 2 are significant, two variables (Unsupported Statements and Cogency Errors) have by far the largest Sum of Squares. One might reasonably ask whether these two variables alone predict the reader holistic score; that issue was examined in a cross validation of another sample. In examining this question, another sample consisting of twenty-nine first and second papers written by different students was drawn at random from

TABLE 2: SYNTHETIC REGRESSION ANALYSIS OF COMPOSITION ATTRIBUTES

$$R^2 = .91$$

ANOVA	DF	S.S.	M.S.	F.	PROB F
Regression	6	264.156	44.026	47.61	0.0001
Error	29	26.816	0.925		
Total	35	290.972			

	B VALUE	STD ERROR	TYPE II ss	F.	PROB F
Intercept	10.783				
Unity	− 0.053	0.017	9.061	9.80	0.0040
Unsupported Statements	− 0.046	0.005	65.750	71.10	0.0001
Cogency	− 0.097	0.010	82.907	82.907	0.0001
Coherence	0.040	0.013	9.195	9.94	0.0037
Vocabvar	− 1.641	0.746	4.475	4.84	0.0359
Subordination	1.845	0.768	5.341	5.78	0.0229

Note: This table is based on the ratios from the data entries of Unsupported Statements, Unity, Cogency, and Coherence errors being multiplied by 100 to avoid decimals in keypunching.

85

the same classes as the first sample. Table 3 shows the cross validation statistics.

Four cross validation tests were conducted. First, the regression equation (Table 2) with six variables was used to predict the reader's holistic score from the student's errors in papers one and two $(r_y,_{\hat{y}} = .87)$. Second, only two variables, unsupported statements and cogency errors, were used for prediction $(r_y,_{\hat{y}} = .90)$. Third, three variables, Unsupported Statements, Unity, and Cogency Errors predicted the holistic scores $(r_y,_{\hat{y}} = .91)$. Fourth, four variables, Unsupported Statements, Subordination, Unity, and Cogency Errors predicted holistic scores $(r_y,_{\hat{y}} = .88)$. These thrid and fourth cross validations may seem unnecessary in that Unsupported Statements and Cogency Errors substantially predicted the readers' holistic scores. Further, the addition of Unity errors does not increase the prediction appreciably, and Subordination actually decreases the prediction. However, both Unity Errors and Subordination have consistently appeared in past studies either as a predictor of or an influence on readers' holistic scores, and these variables appear, again, in a stepwise regression and analysis of papers one and two, as shown in Table 4.

These data can be interpreted in two ways. From a purely practical standpoint, one needs to know only about students' Unsupported Statements and Cogency Errors to reliably predict the readers' assessment of the papers' quality. However, such a practical interpretation does not fit

86

TABLE 3: CROSS VALIDATION STATISTICS

	Paper I + II n = 29		Final Paper n = 36	
	$\overline{X}$	S.D.	$\overline{X}$	S.D.
Unity Errors	0.051	0.11	0.053	0.101
Unsupported Statements	0.694	0.626	0.326	0.401
Cogency Errors	0.270	0.201	0.171	0.181
Subordination	1.87	0.344	1.708	0.238
VOCABVAR	2.03	0.190	2.073	0.264
Coherence Errors	0.098	0.166	0.092	0.158
Y (holistic average)	6.68	3.197	7.47	2.883
Predicted Y (from final paper)	5.203	4.205	not applicable	

$r_{y\ \hat{Y}} = .87 \quad (p < .0001)$ $\hat{Y}$ based on Unity + Unsupported Statements + Subordination + VOCABVAR + Cogency + Coherence

$r_{y\ \hat{Y}} = .90 \quad (p < .0001)$ $\hat{Y}$ based on Unsupported Statements + Cogency

$r_{y\ \hat{Y}} = .91 \quad (p < .0001)$ $\hat{Y}$ based on Unity, Unsupported Statements + Cogency

$r_{y\ \hat{Y}} = .88 \quad (p < .0001)$ $\hat{Y}$ based on Unity, Unsupported Statements, Cogency + Subordination

87

TABLE 4: STEPWISE REGRESSION ANALYSIS

OF PAPERS I + II

$$R^2 = 0.87$$

	DF	Sum of Squares	Mean Squares	F	Prob >
Regression	4	248.476	62.119	39.51	0.000?
ERROR	24	37.730	1.572		
TOTAL	28	286.207			

	B Value	STD Error	Type II SS	F	Prob ⌐ F
Intercept	13.534				
Unity Errors	− 0.057	0.024	8.638	5.49	0.028
Unsupported Statements	− 0.032	0.005	77.223	49.12	0.0001
Cogency Errors	− 0.055	0.013	26.343	16.76	0.0004
Subordination	− 1.537	0.718	7.202	4.58	0.0427

Note: This table is based on the ratios from the date entries of Unsupported Statements, Unity, Cogency, and Coherence errors being multiplied by 100 to avoid decimals in keypunching.

88

present theories of grading, even though those theories have not been empirically grounded. English teachers are not likely to look favorably on a data interpretation indicating that grading, or even the prediction of grading, is simply a bivariate function or a two dimensional mental activity.

Moreover, it would be a mistake, given the limited size of the data, to draw so definite a conclusion about the dimensional nature (space) of the readers' grading. Instead, this study might more reasonably suggest which variables, among the many that have been mentioned and tested, are worth pursuing in future studies. In this respect, four variables consistently show up as predictors: Unsupported Statements, Subordination, and errors in Unity as well as Cogency. These variables are discussed below as they relate to traditional affirmations of writing quality found in composition textbooks, grading scales, and other articles about grading.

Idea development involves a logically reasoned idea with sufficient, relevant, and specific detail. In this study, Unsupported Statements (abstract statements with no supportive detail and with hasty judgements) were inversely related to writing quality (r = -.76), but Unsupported Statements was the single best predictor of the readers' grading. Closely related as a predictor of reader grading were Cogency errors (r = -.75)-- logical fallacies, oversimplifications, contraditions, and erroneous conclusions drawn from detail given in the student's paper. Indeed, Unsupported

89

Statements and Cogency Errors together, according to Goodnight's Max-R procedure, are the two most potent variables in a regression equation predicting reader grading. The R^2 of this equation equaled .816. Further, the cross validation r was .90--again pointing to their importance.

The predominance of these two variables and, thus, of idea development to predict reader grading agrees with others' assertions. Hillocks (1975) points out, from his observations as a teacher, that the single most glaring flaw in student writing from elementary school through college freshman composition is the unsupported statement and the students' lack of specificity. Symes (1972) makes a similar observation. In research studies, the judges of writing quality in Slotnik's (1971) study of computerized grading, as well as the lawyers, teachers, editors, businessmen, and professional writers in Diederich's (1974) study, point to the pre-eminence of the expression of ideas and their development in determining the quality of student writing.

The Organization or Unity of a paper indicates the students' concept of ideas expressed in the paper. If the concept is clear and ordered, the paper has a specific thesis and paragraphs are ordered in a unified manner to explain the thesis. Unity errors in this study are symptomatic of weak concepts. Errors were marked when a paper strayed from the stated topic, when the topic was vaguely stated, or when a paragraph seemed out of order. The appearance of unity errors as a predictor underscores Butler's (1936) findings and reinforces the placement of organization in many writing scales.

Syntactic Maturity may involve subordination of ideas with gerunds, participles, infinitives, and prepositions, but this study focused on the use of clauses as the primary technique of subordination. The variable Subordination was the ratio of clauses to T-units. The other syntactic Variables, T-units and Clauses, although initially included in the regression equation, did not show up as significant predictors. Neither appeared in this study as a very precise indicator of the students' heavy reliance upon the clause as the primary subordination technique. The mean T-unit length (roughly, the typical sentence length) was 1.7 clauses and shows the students' heavy use of clauses more clearly than the ratio of clauses to the paper's words (clause length).

The appearance of Subordination in the regression equation supports the conclusions of Combs (1976, 1977) and O'Hare (1973), who argue for the importance of syntactic maturity in determining the quality of student writing. Combs and O'Hare show that papers exhibiting substantially more subordination are likely to be picked as superior in quality over those that do not show such subordination. In other words, subordination seems to be one basis for teacher evaluation of papers. However, Marzano (1976) points out that syntactic maturity, by itself, does not correlate highly with writing quality; in this study, it was only a moderate predictor of writing quality. Nevertheless, the data in this study support other studies in suggesting subordination as an important dimension, along with other variables, that neither students nor teachers should ignore in appraising good writing.

91

Conclusion

This study provides some useful parameters in considering writing quality. If one extreme in research is to consider only one variable of writing quality, the other extreme is to accept or to judge student themes by too many criteria.

This study provides evidence that perhaps four variables, or qualities, can predict teachers' assessments of student writing quality: Errors in Unity and Cogency, in addition to Unsupported Statements and Subordination. Further, these variables seem to support what some other composition teachers, educators, and researchers have thought as decisive in considering writing quality. While this study does not show that Unsupported Statements, errors in Unity, Cogency or lack of Subordination will cause teachers to down-grade students' themes, it does provide very strong evidence that these elements may be decisive in determining a grade.

References

Braddock, R., Lloyd-Jones, R., and Schoer, L. Research in written
  composition. Champaign, Ill: NCTE, 1963.

Butler, D. C. Frequencies of errors in freshman themes, English Journal,
  1936, 25, 57-60.

Cohen, A. M. Factors accounting for the variance in junior college students'
  composition writing. UCLA, Los Angeles, California, June 1971. ED 060 829.

Combs, W. E. Further effects of sentence-combining practice on writing
  ability, Research in the Teaching of English, 1976, 10, 137-149.

Combs, W. E. Sentence-combining practice: Do gains in judgement of error
  writing quality persist? Journal of Educational Research, 1977, 70,
  318-321.

Diederich, P. B. Cooperative preparation and rating of essay text, English
  Journal, 1967, 56, 573-584.

Diederich, P. B. Measuring growth in English, Urbana, Ill.; NCTE, 1974.

Drucker, P. The surprising seventies, Strategies in prose, W. A. Ferrell
  and N. A. Salerno (Eds.). New York: Holt, Rinehart, and Winston 1973.

Follman, J. C. and Anderson, J. A. An investigation of the reliability of
  five procedures for grading English themes, Research in the Teaching
  of English, 1967, 1, 190-200.

Godshalk, F. I., Swineford, F., and Coffman, W. E. The measurement of
  writing ability. New York: College Entrance Examination Board, 1966.

Hillocks, G. *Observing and writing*. Urbana, Ill.: NCTE, 1975.

Hunt, K. W. Syntactic maturity in school-children and adults, *Monographs of the society of research in child development* (1, Serial No. 134), 1970.

Klein, M. and Grover, B. *An assessment of the effectiveness of symbolic logic in the teaching of composition*. Plymouth, Wis.: Cooperative Educational Service Agency, 1970. (ED 059-209)

Lucas, F. L. What is style? *The critical man: reading, thinking, writing*. A. M. Tibbetts and Charlene Tibbetts (Eds.) Glenview, Ill.: Scott Foresman and Co., 1972, pp. 124-131.

Marzano, R. J. The sentence-combining myth, *English Journal*, 1976, 65, 57-59.

Mellon, J. C. *Transformational sentence combining*, NCTE Research Report, No. 10, Champaign, Ill.: NCTE, 1969.

O'Hare, F. *Sentence combining: improving student writing without formal grammar instruction*. NCTE Research No. 15. Urbana, Ill.: NCTE, 1973.

Orwell, G. Politics and the English language, *Rhetorical considerations: essays for analysis*. H. Brent and W. Lutz (Eds.) Cambridge, Mass.: Winthrop Publishers, Inc., 1974, pp. 1-11.

Page, E. B. and Paulus, D. *The analysis of essays by computer*. U. of Conn., Storrs, Conn., 1968.

Roberts, P. How to say nothing in five hundred words, *Rhetorical Considerations: Essays for Analysis*. H. Brent and W. Lutz (Eds.) Cambridge, Mass.: Winthrop Publishers, Inc., 1974, pp. 21-31.

Slotnik, H. B. An examination of the computer grading of essays.

    Unpublished doctoral dissertation. U. of Ill. at Urbana–Champaign,

    Ill., 1971.

Slotnik, H. B. Writing: National results--writing mechanics. Denver,

    Col.: Educational Commission of the States, Feb. 1972, ED 062-325.

Stoner, D., Beall, L. L., and Anderson, A., A systems approach to the

    teaching of the mechanics of English expression, Research in the teaching

    of English, 1972, 6, 200-211.

Symes, K. M. Frittered away by details: some thoughts on teaching specifics.

    Paper presented at the NCTE Convention, Minneapolis, Minn., 1972,

    ED 077-007.

Thompson, R. F. Predicting writing quality: writing weaknesses that

    dependably predict holistic evaluations of freshman compositions, English

    Studies Collections, 1976, 1, 1-14.