The Relationship between Children's Writing Performance and Standardized Test Scores

Linda Crocker, Linda Lamme, and Toni Ondresik

University of Florida

As American Educators once again focus their attention on the "three R's," the sentiments of many specialists in measurement and language arts have been expressed by McCaig (1977) who stated, "to evaluate achievement in writing, evaluate the writing of children." The inclusion of writing samples by the National Assessment of Educational Progress and the College Entrance Examination Board in their testing programs are two prominent illustrations of the prevailing zeitgeist.

Obtaining reliable estimates of writing ability is an expensive, time-consuming process (Coffman, 1971). A number of studies have demonstrated that writing samples must be obtained from students in multiple modes on multiple occasions and scored by several raters to obtain estimates of students' true abilities (McColly, 1970; Braddock, et al., 1963; Llabre, 1978). Although it may be necessary to evaluate actual student writing samples for certification of mastery or diagnosis of specific difficulties, it still seems advisable to know the extent to which performance on standardized achievement tests is related to actual compositional writing behavior in the classroom. If, for example, there is a substantial relationship between performance in these two areas, then test score information might be used (in conjunction with other information) for preliminary grouping for instruction in writing at the outset of the school year.

The relationship between objective tests and essays has often been explored (e.g., Godshalk, et al., 1966; Klein & Hart, 1968; or Noyes, 1963). These studies have usually involved students at the secondary or college level. Hogan and Mishler (1980) seem to have been the first to examine the relationship for elementary school examinees and in their study, assessment of writing performance was restricted to holistic rating of a single writing sample. The relationship between objective test scores and both mechanistic and holistic qualities of children's compositions remains unknown. Furthermore, the generalizability of such findings for different types of writing assignments has never been demonstrated.

The purposes of this study were to determine:

(1) To what extent is writing performance of children related to scores on a standardized achievement battery?

(2) Which aspects of mechanistic and holistic writing performance are related to performance on objective standardized tests?

(3) Is the relationship similar for different types of writing assignments?

Methodology

Sample

A total of 112 children from four fourth-grade classrooms in a public school in Alachua County, Florida, participated in the study. The student sample was evenly divided between male and female and was comprised of approximately 70 percent white and 30 percent black children.

Instruments

Standardized Test Battery. The Metropolitan Achievement Test (1970) (elementary form) was used in this study. Four specific subtest scores were used: Word Knowledge, Reading Comprehension, Language Arts, and Spelling. The general achievement battery was chosen because data from such tests are more generally available to classroom teachers.

Writing Exercises. Two different types of writing samples were obtained for this study. For the first writing assignment, the children were required to write a fictional short story to explain some phenomenon such as "How the Camel Got His Hump" or "How the Elephant Got His Trunk," etc. Children were permitted to select one of several animals as the topic for the stories. For the second writing assignment, the children were required to prepare a report from a list of facts on the topic of "Bats." Two different types of writing assignments were used so that the generalizability of the relationship between test scores and writing performance could be tested across assignments. These types of assignments were chosen because they typify the kinds of assignments commonly used to teach compositional writing in the classroom.

Scoring the Writing Samples. Writing samples were scored for a total of eight mechanistic and holistic qualities. Mechanistic subscores for each composition included: number of words; number of T-units (thought units); spelling errors; capitalization errors; punctuation errors; mechanical errors (e.g., run-on sentences); and word usage errors. To equate for the fact that different children wrote themes of different lengths, number of T-units and number of errors were adjusted to number per 100 words. In addition, four raters rated each composition for holistic quality using a seven-point scale. The average of

the four ratings was used as the measure of holistic quality. A generalizability coefficient for inter-rater reliability (treating raters and topics as random factors) estimated by analyses of variance was .70. Inter-rater reliabilities between rater pairs ranged from .65 to .77, with a median value of .72. This is comparable to reliabilities reported by Veal and Biesbrock (1971) in their standardized scale for rating essays of students in primary grades.

Data Collection Procedures

The standardized test battery had been administered as a regular part of the spring countywide testing program at the end of third grade for these subjects. In the fall, writing samples were collected for these same students in their fourth-grade classrooms. Instructions for the writing assignment were given by one of the investigators with the teachers in attendance to ensure standard conditions of administration. Children were informed that the writing samples would be graded and returned to them by their teachers so that they would be motivated to perform as well as they would for any other writing assignment. One week elapsed between the two writing assignments to avoid effects of fatigue and boredom.

Analysis and Results

The first question addressed in this study dealt with the extent to which standardized test scores were related to writing performance. Both multivariate and univariate analysis procedures were used to explore this area of inquiry. First, a multivariate procedure, canonical correlation, was used to determine the amount of common variance between the set of four standardized test scores and the set of eight

measures of writing proficiency. This seemed especially appropriate because most teachers consider "writing proficiency" to be a combination of mechanistic skills and holistic quality which occur simultaneously in a given sample of student written work. Separate analyses were conducted for the fiction and factual report compositions to illustrate the generalizability of findings across different types of written assignments. Results of the canonical variate analyses are presented in Tables 1 and 2. Results of these analyses can be summarized as follows:

(1) For the fiction writing assignment, one canonical variate was extracted which accounted for a significant proportion

Table 1

Results of Canonical Variate Analysis between Standardized Test Scores and Measures of Writing Proficiency in Fiction (N = 70)

Variables	Standardized Coefficients First Canonical Variate	Canonical Corr.
Standardized Tests		
1. Word Knowledge	.25	.70*
2. Reading Comprehension	.64	
3. Language Arts	-.23	
4. Spelling	.37	
Measures of Writing Proficiency		
1. Number of Words	-.44	
2. Number of T-units	.17	
3. Spelling Errors	-.29	
4. Capitalization Errors	.04	
5. Punctuation Errors	-.14	
6. Mechanical Errors	-.34	
7. Word Usage Errors	-.14	
8. Holistic Rating	.88	

*Significant at $\alpha = .05$
(Wilk's $\Lambda = .3807$, $X^2 = 60.36$, with 32 df, $p \leq 002$)

of variance in the two sets of variables. The canonical correlation of .70 was significant at $\alpha = .05$. The second canonical variate extracted failed to account for a significant portion of common variation in performance on these two sets of variables. This indicates that, even though the standardized test battery and the measures of writing proficiency might have complex underlying factorial structures if considered separately, the joint variation for these two sets of variables can be adequately attributed to a single source.

Table 2

Results of Canonical Variate Analysis between Standardized Test Scores and Measures of Writing Proficiency in Factual Report (N = 68)

Variables	Standardized Coefficients First Canonical Variate	Canonical Corr.
Standardized Tests		
1. Word Knowledge	-.47	.64
2. Reading Comprehension	1.14	
3. Language Arts	.09	
4. Spelling	.21	
Measures of Writing Proficiency		
1. Number of Words	-.20	
2. Number of T-units	-.02	
3. Spelling Errors	-.12	
4. Capitalization Errors	-.44	
5. Punctuation Errors	-.22	
6. Mechanical Errors	-.43	
7. Word Usage Errors	-.22	
8. Holistic Rating	.64	

(Wilk's $\Lambda = .4833$, $X^2 = 43.99$, with 32 df, $p \leq .07$)

(2) The standardized test variable with the greatest weight on the significant canonical variate was Reading Comprehension. The writing proficiency variable with the greatest weight was the Holistic Rating. Due to the relatively small size of the standardized coefficients for the remaining seven writing variables, it seemed appropriate to focus on the holistic rating as representative of the set of writing variables in a follow-up multiple regression analysis.

(3) For the factual report assignment, the canonical correlation (.64) between the two sets of variables fell short of the level of significance ($p = .07$). In view of this, examination of the variable weights is questionable (although again the major weighting among the standardized test variables was for Reading Comprehension and on writing proficiency, the Holistic Rating received the major weighting).

From the canonical analysis, the holistic rating was identified as the most important of the eight measures of writing proficiency. To determine the degree to which the four standardized test scores were related to holistic writing the four test scores were regressed on the single criterion of the holistic rating for the fiction assignment using the standard regression solution described by Nie, et al. (1975). (This follow-up analysis was performed only for the fiction assignment because the canonical correlation for the factual report assignment was not statistically significant.) The correlation matrix for the five variables entered into the regression analysis is presented in Table 3. All correlations are significant at alpha = .05.

Table 3

Correlation Matrix for Standardized
Test Scores and Holistic Rating

Variables	1	2	3	4	5
1. Word Knowledge		.84	.58	.59	.45
2. Reading Comprehension			.58	.59	.52
3. Language Arts				.45	.30
4. Spelling					.43
5. Holistic Rating					--

Results of the multiple regression analysis are presented in Table 4. The multiple R of .55 was obtained between standardized test scores and holistic writing score and the R^2 (.30) was significant at $\alpha = .05$. Examination of the standardized regression coefficients for the four tests and their corresponding F-ratios indicated that Reading Comprehension made the significant contribution to the multiple correlation. The failure of the other three subtests to receive significant weights in the regression equation does not mean that these variables were unrelated to holistic writing proficiency. It simply indicates that they did not share additional unique variance with the criterion beyond that parceled out with the Reading Comprehension subtest.

Discussion

For the fourth graders who participated in this study, there was a relationship between standardized achievement scores and writing proficiency. The somewhat unexpected finding was that the Reading

Table 4

Results of Multiple Regression of Standardized Test Scores on Holistic Rating

Multiple R = .55, R^2 = .30

F with 4,65 d.f. = 6.972*

Variable	Standardized Regression Coeff.	Standard Error of Regression Coeff.	F
Word Knowledge	-.0034	.0004	0.000
Reading Comprehension	.5005	.057	4.989*
Language Arts	-.1281	.0100	0.776
Spelling	.1845	.0138	1.898

*Significant at α = .05

Comprehension subtest surpassed the Language Arts subtest as a predictor of writing performance. This seems more reasonable, however, when the tasks involved in each subtest are analyzed. In the Reading Comprehension test, examinees are required to read a paragraph and identify main ideas or draw inferences from the passage. In holistic rating of a composition, evaluation is based on the writer's ability to organize and express his thoughts around main ideas in logical sequence. In short, reading comprehension requires recognition of main ideas in written passages, while effective composition requires production of main ideas and their elaboration within the context of written passages. Thus the student who has failed to acquire the concept of what a "main idea" is or how it functions, probably can neither read with understanding nor write with clarity. By contrast, in the Language Arts subtest,

the examinee must read sentences and identify errors in punctuation, capitalization, etc. Thus performance on this subtest may be more closely related to mechanistic scores on a composition than to the holistic rating of composition quality. It should also be noted that Hogan and Mishler (1980), using a more recent (1978) edition of the Metropolitan Achievement Test, reported slightly higher correlation between holistic writing and the language arts subtest for third graders in their study.

When test scores were related to holistic ratings on factual reports, the relationship between test scores and writing performance was not significant at the .05 level. The failure of this second canonical correlation to achieve statistical significance underscores the need to assess student writing in a variety of modes to determine generalizability of findings. From a less technical viewpoint, however, the relatively high canonical r-value (just missing the .05 level of significance) and the similarities in the loading patterns on the first canonical variate for the fiction and factual writing samples seem to support, rather than contradict, the first finding.

In any event, the results of this study must be interpreted with caution. Although the study offers some evidence of overlap between the abilities that underlie standardized test performance and the abilities that underlie writing performance, the strength of the relationship observed was insufficient to suggest that standardized tests can be used to replace actual measures of writing. Furthermore, any attempt to establish predictive equations using the regression weights derived for the sample in this study should be especially discouraged since no cross-validation has been conducted.

What, then, if any, are the implications of this study for assessment practices at the classroom level? Viewing the outcomes of this study within the broader context of research in the area of writing assessment, the following recommendations are offered:

(1) When grouping students for instruction in compositional writing, teachers may wish to use both standardized test scores and actual writing samples as corroborating pieces of information.

(2) More than one type of writing sample should be considered.

(3) Teachers should not consider Language Arts or Spelling subtests as the only subtests related to writing performance. They should also attend to students' scores in Reading Comprehension, since this has been shown to have significant relationship to writing.

(4) Teachers should expect some substantial discrepancies to occur between test scores and writing measures. (In this study more than half of the students displayed performance discrepancies of at least half a standard deviation.)

Other considerations in use of student writing samples for placement which were not directly addressed in this study, are the effects of handwriting on the ratings of student essays (Chase, 1968 and 1979) and the possibility that knowledge of test scores could create expectancies about student performance that are reflected in essay ratings (Chase, 1979). It is probably advisable to control for both of these factors when student writing samples are used for placement or evaluation.

References

Beal, L. R., & Biesbrock, E. F. Primary essay tests. Journal of Educational Measurement, 1971, 8, 45-46.

Braddock, R., Lloyd-Jones, R., & Schoer, L. Research in written composition. Champaign, Ill.: National Council of Teachers of English, 1963.

Chase, C. I. The impact of achievement expectations and handwriting quality on scoring essay tests. Journal of Educational Measurement, 1979, 16, 39-42.

Chase, C. I. The impact of some obvious variables on essay test scores. Journal of Educational Measurement, 1968, 5, 315-318.

Coffman, W. E. Essay examinations. In R. L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council on Education, 1971.

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

Hogan, T. P., & Mishler, C. Relationships between essay tests and objective tests of language skills for elementary students. Journal of Educational Measurement, 1980, 17, 219-227.

Klein, S. P., & Hart, F. M. Chance and systematic factors affecting essay grades. Journal of Educational Measurement, 1968, 5, 197-206.

Llabre, M. M. An application of generalizability theory to the assessment of writing ability. Unpublished doctoral dissertation, University of Florida, Gainesville, Fla., 1978.

McCaig, R. A. What your director of instruction needs to know about standardized English tests. Language Arts, 1977, 54, 491-495.

McColly, W. What does educational research say about the judging of writing ability? Journal of Educational Research, 1970, 64, 148-156.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. Statistical package for the social sciences. New York: McGraw-Hill, 1975.

Noyes, E. S. Essay and objective tests in English. College Board Review, 1963, 49, 7-10.