DOCUMENT RESUME

ED 052 206 TE 002 493

AUTHOR	Whalen, Thomas E.
TITLE	Assessment of Language Ability by Computer.
PUB DATE	Apr 71
NOTE	18p.; Paper presented to the California Educational Research Association (April 1971)
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	*Composition Skills (Literary), *Computer Programs, *Essays, Junior High School Students, Language Ability, Models, *Predictor Variables, *Writing Skills
IDENTIFIERS	PEG, Project Essay Grade

ABSTRACT

A study was made of the capability of machine scoring of essays to determine a student's proficiency in English mechanics. A sample of 71 seventh grade essays were entered into a computer via punched cards and were processed using a modified version of the PEGFOR program. Prediction models, subtest models, and restricted models were tested for predictive efficiency. The most reliable model constructed was composed of seven variables: (1) number of capitalization errors, (2) standard deviation of sentence length, (3) standard deviation of word length, (4) number of connectives, (5) occurrence of "then", (6) average sentence length, and (7) number of usage errors. When used to predict overall essay grades, using regression weights derived from the mechanics criterion, the correlation between actual and predicted scores for 24 essays was .60, indicating a strong relationship between mechanical proficiency and overall writing ability at the seventh grade level. It is believed that computer scoring of essays may prove to be an important tool in the individualization of composition teaching. (DB)

ASSESSMENT OF LANGUAGE ABILITY BY COMPUTER

A Paper Presented to

The California Educational Research Association

April, 1971

by

Thomas E. Whalen

California State College, Hayward

ASSESSMENT OF LANGUAGE ABILITY BY COMPUTER

Previous research (Page and Paulus, 1968) has shown the efficacy of analyzing student essays by computer. Using an actuarial approach, the researchers were able to obtain a multiple correlation of .72 between their thirty computer-derived predictors and an average of human judgments for five writing traits--content, organization, style, mechanics, and creativity. In a related study by Janzen (1968), twenty-two linguistic variables similar to those used by Page and Paulus were utilized as criteria of writing ability. The author concluded that "this verifies Page's findings that these variables are useful as 'proxes' for a measure of a student's ability to write English compositions" (p. 47).

One of the major problems confronting the Project Essay Grade (PEG) researchers was the formidable task of human judgment of the essays. Although thirty-two teachers participated, each was required to read and grade sixty-four essays on five writing traits. The average time allowed for the multiple-trait judging was 3½ minutes (Page and Paulus, p. 76). This procedure generated high correlations between traits, but such correlations were thought to be due, in part, to a "halo" effect. It was noted, however, that mechanics seemed to be least effected by halo. It had the largest standard deviation of any trait (p. 77) and the highest human-group reliability (p. 103). Despite this fact, mechanics was the second most difficult trait to predict.

The purpose of this study was to seek answers to the following questions: (1) Can the prediction of mechanical accuracy be improved through the use of additional variables and by achieving a more objective and reliable measure of mechanics? and (2) Can computer evaluation of essays predict with any degree of success student scores on a standardized test of English mechanics including capitalization, punctuation, word usage, and spelling?

PROCEDURES

The Sample

A sample of seventy-one seventh grade essays was used in the study. Shortly after reading The Adventures of Tom Sawyer, students from three "average" classes were asked to write a letter to a friend about the book. Specifically, students were directed to write a 200-word composition on any aspect of Tom Sawyer they chose. The students were given thirty minutes writing time and were not allowed to correct or rewrite their papers.

The California Language Test, Junior High Level, was administered to the same students shortly after their writing of the essays. This test includes three separate sections, two of which were used for this study--Mechanics of English and Spelling.

Predictor Variables

The essays were entered as computer input data via punched cards to a modified version of the PEGFOR computer program (Whalen, 1970). Frequency counts and other measures were taken for seventeen variables previously used by Page and Paulus: number of (1) paragraphs, (2) parentheses, (3) commas, (4) colons, (5) semicolons, (6) quotation marks, (7) question marks, (8) prepositions, (9) connective words, (10) spelling errors, (11) relative pronouns, (12) subordinating conjunctions, and (13) words on the Dale List. Means and standard deviations were calculated for word and sentence length. In addition, nine "new" predictors were generated by the PEGFOR program. These variables were selected as potential predictors of both mechanical proficiency and overall writing ability. They included the (1) type-token ratio, occurrences of (2) so, (3) and, (4) when, (5) then, (6) forms of the verb to be, and number of (7) capital letters,

(8) capitalization errors, (9) usage errors.

Rationales for the first set of variables are set forth by the PEG researchers (Page and Paulus, 1968). A brief discussion concerning the inclusion of the second set follows:

(1) *Type-token ratio*. Edmundsen (1967) defined the type-token ratio as the number of word types divided by the number of word tokens, i.e., the number of uniquely different words divided by the total words in the text. It was hypothesized that a relatively high ratio would correlate positively with good writing.

(2) Occurrences of *so*. The word *so* was believed to be associated with run-on sentences.

(3) Occurrences of *and* and (4) *when*. Hunt (1964) provided evidence that occurrences of *and* and *when* are strongly related to writing maturity. In samples of student writing he found significantly decreasing occurrence frequencies for both of these words from the fourth to the twelfth grade levels.

(5) Occurrences of *then*. This word is often found in the same context as *so*. It was included for similar reasons and was hypothesized to relate negatively with essay quality for seventh graders.

(6) Forms of the verb *to be*. Many stylists including Tanner (1968) have suggested that good writing is marked by the absence of the verb *to be* and its forms. A major reason for this is that all passive constructions require a form of *to be*. A preponderence of passives is thought to be indicative of a weak, imprecise style.

(7) Number of capital letters. Since most errors in capitalization are errors of ommission, it was hypothesized that a high number of capital letters would indicate freedom from such errors.

(8) Number of capitalization errors and (9) usage errors. In order to obtain counts for these variables, two special "dictionaries" were

constructed by this investigator and included as data to the PEGFOR program. The first was a dictionary of proper nouns which included all the proper nouns from Tom Sawyer and other nouns such as the days of the week, holidays, etc.--177 words in all. Although this predictor must be considered specific to the content of the essays written, its inclusion was made to determine the usefullness of such a predictor in cases where prediction is made for a very limited and perhaps, standardized test of writing ability. At present it is not feasible to include a dictionary of all possible English proper nouns in the PEGFOR program.

A second dictionary of over 500 one- and two-word usage errors was constructed and entered as input to sub-routine PHRASE of the revised PEGFOR program. Entrees for this dictionary were taken from a number of sources including English grammar and usage texts, standardized tests, and other student essays. Examples of usage errors are knowed, drowned, has went, could of, and ain't got, etc.

Criterion Variables

The first dependent variable, mechanical proficiency, was defined in terms of the total number of mechanical errors in a student's essay. In order to provide an objective measure of this criterion, a procedure based on previous research (Whalen, 1969) was used. Three raters, including this investigator and two graduate assistants, were utilized to determine error counts for the essays. The two graduate students were both prospective teachers with degrees in English.

Due to the extensive nature of the rating task, each rater judged only a portion of the essays. During the early phases the three raters worked together on several essays to achieve agreement on scoring procedures. A total of sixteen error categories was established. These categories served as a guide to the individual raters through the remainder of the rating period. Table 1 shows essay error totals for an initial sample of seventy-seven essays (six of these were dropped due to their short lengths). It can be seen that spelling and punctuation errors accounted for more than half of all errors. Capitalization errors and run-on sentences also accounted for a substantial portion. These four categories represented more than three-fourths of all errors committed.

TABLE I

Essay Error Totals

Error Type	Type Total
Spelling	359
Punctuation	340
Capitalization	126
Run-on	107
Wrong Word[1]	71
Word Omission	64
Verb Tense[2]	37
Extra Word	22
Fragment	20
Verb Ending[3]	19
Agreement[4]	15
Pronoun Reference[5]	13
Indentation	12
Awkward[6]	11
Illegible	7
Pronoun Case[7]	6
Total Errors	1,239

1. Know, Aunt Polly, I haven't been swimming.

2. They runned from Injun Joe.

3. Becky felt better and stop crying.

4. Life without adventures were dull.

5. They were looking for dead bodies and dug it up.

6. Huck wasn't well-liked at all means by the moms.

7. Becky and him had a fight.

The second criterion for prediction was termed "language ability" because it was measured by the California Language Test, Form W (1957). However, the term is not intended to connote the broad spectrum of verbal abilities usually associated with language ability. This criterion, from a conceptual standpoint, is actually the same as "mechanical proficiency" as discussed above.

Test 5 of the California Language Test, Mechanics of English (1957), consists or three sections--Capitalization, Punctuation, and Word Usage. According to the authors, these three sections sample twenty-two different elements of the mechanics of English and provide an aid in diagnosing the specific difficulties encountered by students in this area. There are thirty questions on capitalization, twenty-nine questions on punctuation, and thirty questions on word usage.

Test 6, Spelling (1957), consists of thirty items in which students must identify misspelled words from groups of four words for each item. For the present study, total error scores on the four tests were used as the criterion of language ability. In addition, an attempt was made to predict scores on the separate subtests, as well. Reliability co-efficients for the test were reported as follows: .92 for Mechanics of English, .83 for Spelling, and .93 for Total Language.

Statistical Methods

Data generated by the PEGFOR program were used to calculate linear regression equations for all criterion variables. In order to control for essay length, frequency counts were made for only the first 200 words of each essay. A step-wise regression program (IBM Scientific Subroutine Package, Version III, 1968) was used to calculate the regression coefficients. The prediction equations were formed by using scores from two-thirds (forty-seven) of the sample essays. The prediction models were then cross-calidated by applying them to the remaining one-third (twenty-four) of the essays.

In addition to constructing prediction models containing all twenty-six predictors variables, an attempt was made to isolate a subset of the most "potent" predictors for each criterion. These subsets, or restricted models, were tested to determine their predictive efficiency.

RESULTS

Full Mechanics Model

Results for the prediction of mechanical proficiency are presented in Table 2. Since the criterion was defined in terms of error scores, it is necessary to interpret correlations with the criterion accordingly. Capitalization errors proved to be an important measure of mechanics for the sample data. This variable correlated highly with total mechanics errors and was selected first by the computer algorithm. It should be recalled that the computer-derived frequency counts for capitalization errors were based on a dictionary of proper nouns taken from the book Tom Sawyer, and were not based on a universal proper noun list. Even so, it is well worth knowing that such a variable can be utilized quite effectively in situations calling for a restricted topic assignment.

TABLE 2

STEPWISE MULTIPLE REGRESSION: MECHANICS

Step	Variable	r	b-wt	SE	t-Value	Mult-R	SE
1.	Cap. errors	.54	1.79	1.37	1.30	.537	12.10
2.	S.D. sent. lgth.	.44	0.42	0.95	0.44	.693	10.46
3.	Av. wd. lgth.	-.50	1.39	39.10	0.03	.732	10.00
4.	No. of caps	-.18	0.41	0.54	0.75	.747	9.87
5.	Quotation mks.	-.16	0.23	1.84	0.12	.770	9.58
6.	Type-token	.15	153.93	105.99	1.45	.786	9.40
7.	Parentheses	-.09	-5.82	5.16	-1.12	.803	9.19
8.	Av. sent. lgth.	.37	0.55	1.14	0.48	.813	9.09
9.	Prepositions	-.37	-0.55	0.50	-1.10	.821	9.04
10.	"Then"	.23	2.41	2.44	0.98	.824	9.08
11.	Usage	.19	-5.23	2.68	-1.86	.826	9.16
12.	Colons	-.05	-34.95	25.83	-1.35	.829	9.23
13.	S.D. wd. lgth.	-.47	-42.88	29.23	-1.46	.832	9.29
14.	Subord. conj.	-.10	-1.65	1.39	-1.19	.837	9.31
15.	Spelling	.14	3.73	2.85	1.30	.841	9.34
16.	Connectives	-.19	-4.12	3.56	-1.15	.846	9.37
17.	Commas	.11	-0.50	0.64	-0.78	.852	9.35
18.	Paragraphs	-.02	0.72	0.94	0.76	.857	9.37
19.	"When"	-.16	1.27	2.17	0.58	.859	9.47
20.	Semicolns	-.02	-3.07	7.70	-0.39	.861	9.61
21.	Question mks.	.17	3.04	13.79	0.22	.861	9.79
22.	"To be"	.04	-0.22	0.99	-0.22	.861	9.98
23.	"And"	-.21	-0.11	0.83	-0.13	.861	10.19
24.	Rel. pronouns	.09	0.16	1.30	0.12	.861	10.42
25.	"So"	-.17	-0.02	1.94	-0.01	.861	10.66
26.	Dale list	-.04	0.00	0.11	0.01	.861	10.93

Intercept Constant = 16.38
F Mult-R = 2.214 (at step 26)

Standard deviation of sentence length was selected second in the stepwise program. The bivariate correlation of .44 lent support to the hypothesis that a wide variation of sentence lengths, most probably due to the more frequent use of run-ons, is associated negatively with mechanical proficiency at the seventh grade level. Other important predictors included average word length, the number of capital letters, quotation marks, parentheses, and prepositions.

Three new predictors in addition to capitalization errors were included among the first eleven variables. The type-token ratio was selected sixth in the stepwise procedure; occurrences of <u>then</u> and usage errors were selected tenth and eleventh, respectively. The mult-R for mechanics was .86.

Language Ability

The prediction of standardized test scores of English Mechanics from a single sample essay was considered intuitively to be a most difficult task. Table 3 shows much better results than anticipated for this model. The first variable selected was average word length with a correlation of .50 with the criterion. Considering that word length has no direct relationship to mechanical proficiency, such a high relationship provides evidence of the robustness of this variable as a general predictor. Variable two in the model, common words on the Dale List, is an example of a suppressor variable at work. Though its correlation with the criterion was fairly low, it was highly related to average word length (-.45). The effect was to partial out a large portion of the residual error variance of average word length and thus make its relationship with language ability stronger.

TABLE 3

STEPWISE MULTIPLE REGRESSION: LANGUAGE ABILITY

Step	Variable	r	b-wt	SE	t-Value	Mult-R	SE
1.	Av. wd. lgth.	-.50	-68.86	38.42	-1.79	.504	13.09
2.	Dale list	-.18	-0.18	0.11	-1.67	.684	11.18
3.	Spelling	.24	0.72	2.80	0.26	.724	10.70
4.	Colons	-.06	24.16	25.38	-0.95	.750	10.37
5.	Subord. conj.	-.28	-1.62	1.36	-1.18	.772	10.09
6.	S.D. sent. lgth.	.37	-0.01	0.93	-0.02	.793	9.80
7.	"Then"	-.11	1.77	2.39	0.74	.800	9.76
8.	Cap. errors	.36	1.08	1.35	0.80	.808	9.72
9.	Rel. pronouns	-.14	-.83	1.28	0.65	.816	9.67
10.	Quotations	-.28	-1.27	1.81	-0.70	.824	9.59
11.	"When"	.35	1.22	2.13	0.57	.828	9.63
12.	Type-token	.21	231.74	104.13	2.22*	.833	9.65
13.	No. of caps	-.24	1.12	0.53	2.09*	.855	9.17
14.	"So"	.17	2.47	1.90	1.29	.864	9.05
15.	"To be"	-.09	0.80	0.98	0.82	.869	9.02
16.	Connectives	-.08	1.66	3.50	0.47	.873	9.07
17.	Usage	.40	1.33	2.75	0.48	.876	9.10
18.	Question mks.	-.11	-5.85	13.55	-0.43	.878	9.20
19.	Parentheses	.09	-2.18	5.07	-0.43	.879	9.32
20.	Prepositions	-.37	0.11	0.49	0.23	.880	9.46
21.	"And"	-.05	-0.34	0.82	-0.42	.881	9.62
22.	Av. sent. lgth.	.35	0.26	1.12	0.23	.881	9.81
23.	Semicolons	-.05	-0.87	7.56	-0.11	.881	10.02
14.	S.D. wd. lgth.	-.47	-3.62	28.72	-0.12	.881	10.24
15.	Commas	-.09	-0.05	0.63	-0.09	.881	10.48
16.	Paragraphs	-.17	-0.07	0.93	-0.08	.882	10.73

Intercept Constant = 185.73

F. Mult-R = 2.681 (at step 26)

* Significant at .05 level

Other predictors exhibiting a relatively high relationship with the criterion were standard deviation of sentence length, capitalization errors, occurrences of when, usage errors, number of prepositions, average sentence length, and standard deviation of word length. All of these relationships were in accordance with preliminary hypothese. The mult-R for the language ability criterion was .88.

Subtest Models

It is generally known that part scores, even on standardized tests, are frequently much less reliable than total test scores derived from the sum of the parts. This is basically a statistical phenomenon and is due to the relatively fewer number of responses measured by the subtests. The subtests of the California Language Test--capitalization, punctuation, usage, and spelling--contained thirty, twenty-nine, thirty, and thirty items, respectively. The authors of the test did not report reliability cofficients for the first three of these tests. The reliability of the spelling test was .83.

Considering the relative instability of these criterion measures, it was somewhat encouraging that all of them were predicted with considerable accuracy. The multiple correlations for the four subtests were .88, .87, .79, .84. Table 4 is a summary of the raw and corrected mult-R's for the six models. An indication of the important relationship between sample size and number of predictors was demonstrated by the considerable shrinkage calculated for the usage criterion. For a sample of forty-seven with twenty-six predictors, minor fluctuations in the mult-R are accompanied by considerable changes in shrinkage as calculated by the Wherry formula.

TABLE 4

SUMMARY OF MULTIPLE CORRELATIONS FOR ALL FULL MODELS

Criteria	Mult-R	Shrunk*	Rel.	Atten.**
Mechanics	.86	.63	–	–
Language Ability	.88	.69	.93	.72
Cal. Capitalization	.88	.69	–	–
Cal. Punctuation	.87	.67	–	–
Cal. Usage	.79	.36	–	–
Cal. Spelling	.84	.57	.83	.63

*Calculated by Wherry formula. N=47

**Mult-R's were corrected for attenuation only for those criteria where reliability coefficients were available.

Restricted Models

Because this study emphasized the development of efficient prediction models, an attempt was made to reduce the number of variables to the most parsimonious set of predictors possible. A limited subset of predictors was selected on the basis of their relative stability, their frequency of occurrence, and the magnitude of their correlations with the criteria. Tables 5 and 6 show the composition of the two restricted models for mechanics and language ability. It is important to note that, although the mult-R's for these models were lower then for the full models, the F-values were considerably higher.

TABLE 5

Restricted ~~FORCED~~ MECHANICS MODEL

Variable	b-Wt	t-Value	Mult-R	F-Value
1. Cap. errors	3.85	4.05**	.54	18.21
2. S.D. sent. lgth.	1.48	2.87**	.69	20.28
3. S.D. wd. lgth.	-27.89	-2.39*	.72	15.36
4. Connectives	-3.17	-1.58	.73	12.16
5. "Then"	2.19	1,89	.75	10.23
6. Av. sent. lgth.	-0.93	-1.79	.76	9.17
7. Usage errors	-1.79	-1.31	.77	8.25**
Intercept	71.30			

* Significant at .05 level
** Significant at .01 level

TABLE 6

Restricted ~~FORCED~~ LANGUAGE ABILITY MODEL

Variable	b-Wt	t-Value	Mult-R	F-Value
1. S.D. wd. lgth.	-16.10	-1.18	.47	12.67
2. Subord. Conj.	-1.91	-2.62*	.59	11.89
3. Cap. errors	2.94	2.71**	.64	10.29
4. "Then"	2.88	2.24*	.68	9.20
5. S.D. of sent.lgth.	0.35	1.25	.70	8.05
6. "So"	2.35	1.42	.72	7:11
7. Usage errors	1.50	0.98	.73	6.23**
Intercept	57.21			

* Significant at .05 level
** Significant at .01 level

The results of cross-validation of these and the other restricted models are shown in Table 7. Correlations between predicted and actual scores for mechanics and language ability were .68 and .60, respectively. These coefficients are both significant at the .01 level of confidence. Less success was noted for the California subtest models. However, these coefficients were all significant at the .05 level.

TABLE 7

CROSS VALIDATION OF FORCED PREDICTION MODELS

Criteria	Mult-R	Shrunk.	Atten.
Mechanics	.77	.68**	–
Language Ability	.73	.58	.60**
Cal. Capitalization	.57	.35*	–
Cal. Punctuation	.63	.45*	–
Cal. Usage	.56	.45*	–
Cal. Spelling	.66	.36	.40*

*Significant at .05 level for one-tailed test.

**Significant at .01 level for one-tailed test.

SUMMARY AND CONCLUSIONS

The final model constructed to predict mechanical proficiency was composed of seven variables: (1) number of capitalization errors, (2) standard deviation of sentence length, (3) standard deviation of word length, (4) number of connectives, (5) occurrence of <u>then</u>, (6) average sentence length, and (7) number of usage errors. The variables in this model had a multiple correlation of .77 with the criterion. This coefficient shrank to .68 after empirical cross-validation.

These results indicated that the mechanics model was the most reliable of all the models in this study. This 7-variable model compared favorably with a 30-variable model constructed by Page and Paulus (1968). Those investigators reported a mult-R of .69 (adjusted by the Wherry formula) for the prediction of mechanics. (1968, p. 103).

In order to determine the general utility of the mechanics model, an attempt was made to predict overall essay grades (as determined by a panel of judges) by using the regression weights derived from the mechanics criterion. Surprisingly, the correlation between actual and predicted scores for twenty-four essays was .60, indicating a strong relationship between mechanical proficiency and overall writing ability at the seventh grade level.

Several of the new variables including number of capitalization errors, occurrences of *then*, and number of usage errors were important contributors to the prediction of mechanical proficiency. Other new predictors such as number of capital letters, occurrences of *and, when*, and the forms of the verb *to be* were less successful. Occurrences frequencies for capital letters and the word *and* were erratic across essays. Perhaps with a substantially longer sample of text, their use might be more profitable.

The machine prediction of language ability as measured by the California Language Test was, indeed, a success. However, accurate prediction of four separate dimensions of the test was less successful. This was due, in part, to the lower reliability of the subtest scores. One variable which was expected to contribute strongly to this and the mechanics model was number of spelling errors. However, its relationship with both criteria was not especially strong. Although there were more than four spelling errors on the average in each essay, the computer was able to detect less than one of them. This would suggest that the dictionary of misspelled words should be augmented to include more words commonly misspelled by less sophisticated writers. Another dimension which was not properly represented in the model was punctuation. None of the punctuation variables appeared to contribute substantially toward predicting the criterion. If additional predictors can be found which adequately measure spelling and punctuation ability, the prediction of language ability should certainly be improved.

In conclusion, it appears that machine scoring of essays for purposes of determing a student's level of proficiency in English mechanics and usage is worthy of further attention. Considerable concern has been expressed by language teachers in recent years that students' writing ability be assessed by direct means rather than through the use of objective tests. Improvements in computer hardware and software features could make this feasible.

Computer scoring of essays may ultimately be an important tool in the individualization of composition-teaching. At present, most English teachers are heavily overburdened and simply find too little time for the assignment and correction of many essays. Consequently, students suffer from lack of writing practice. With the advent of computer time-sharing and "conversational" teletype terminals, it is already possible for students to communicate directly with a computer from the classroom. If a fast and reliable procedure for evaluating writing can be developed, a resurgence and improvement in essay-writing is surely forthcoming.

REFERENCES

California Language Test, Junior High Level, Form W. Del Monte Research Park, Montery, California: California Test Bureau, 1963.

Darlington, R.B. Multiple regression in psychological research and practice. *Psychological Bulletin,* 1968, 69 (3), 161-182

Diederich, P., French, J., and Carlton, S. Factors in the judgment of writing ability. *ETS Research Bulletin, No. 15.* Princeton: Educational Testing Service, 1961.

Edmundson, H.P. Mathematical models in linguistics and language processing. In Borko, H. (Ed.) *Automated language processing.* New York: Wiley and Sons, Inc., 1967. pp. 33-96.

Godshalk, F., Swineford, F., and Coffman, W. *The measurement of writing ability.* Princeton: Educational Testing Service, 1966.

Hunt, K.W. Differences in grammatical structures written at three grade levels. *USOE Cooperative Research Project 1998.* Tallahassee, Fla.: Florida State University, 1964.

Janzen, H.L. *A study of written language ability.* Unpublished Master's thesis. The University of Calgary, Alberta, Canada, 1968.

Mosier, C.I. The need and means of cross-validation. *Educational and Psychological Measurement*, 1951, *11*, 5-11.

Page, E.B., and Paulus, D.H. The analysis of essays by computer. *USOE Project Number 6-1318*, April, 1968.

Tanner, Bernard R. The writers paradox. *English Journal*, 1968, *57* (6), 857-865.

Whalen, T.E. Total English equals writing competence. *Research in the teaching of English*, 1969, *3* (1), 52-61.

Whalen, T.E. A comparison of language factors in primary readers. *The Reading Teacher*, 1970, *23* (6), 563-570.

Whalen, T.E. A computer analysis of mechanical proficiency and overall quality in seventh grade English essays. Unpublished doctoral thesis, University of Connecticut, 1970.

Wherry, R.J. Comparison of cross-validation with statistical inference of betas and multiple R from a single sample. *Educational and Psychological Measurement*, 1951, *11*, 23-28.