1977

Determining the Internal Consistency of English Compositions Using Selected Criteria.

James Howard Wynn

Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_disstheses

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_disstheses/3090

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.
WyNN, James Howard, 1949-
DETERMINING THE INTERNAL CONSISTENCY
OF ENGLISH COMPOSITIONS USING SELECTED
CRITERIA.
The Louisiana State University and
Agricultural and Mechanical College,
Ph.D., 1977
Education, language and languages

Xerox University Microfilms, Ann Arbor, Michigan 48106

Copyright © 1977 by James Howard WyNN
DETERMINING THE INTERNAL CONSISTENCY
OF ENGLISH COMPOSITIONS USING
SELECTED CRITERIA

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in
The Department of Education

by
James Howard Wynn
B.A., Louisiana Tech University, 1970
M.A., Louisiana Tech University, 1973
May, 1977
ACKNOWLEDGMENTS

The author wishes to express his appreciation to the persons who aided in this study. Dr. L. L. Kilgore served as major professor and offered many valuable suggestions on the style and format of the manuscript. Drs. Sam Adams, Nicholas Canaday, J. W. Finnberg, J. J. Parker, W. M. Smith, and B. M. Strawitz were committee members and advisors during the study. Special acknowledgment is also due Dr. Prentiss Schilling who provided the author with suggestions in choosing an efficient statistical procedure.

The writer wishes to pay special tribute to his wife, Gloria Higgins Wynn, who encouraged him throughout the study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>x</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>PURPOSE OF THE STUDY</td>
<td>1</td>
</tr>
<tr>
<td>NEED FOR THE STUDY</td>
<td>1</td>
</tr>
<tr>
<td>THEORETICAL BASIS</td>
<td>2</td>
</tr>
<tr>
<td>SCOPE OF THE STUDY</td>
<td>2</td>
</tr>
<tr>
<td>DEFINITIONS</td>
<td>3</td>
</tr>
<tr>
<td>2. REVIEW OF RELATED LITERATURE</td>
<td>4</td>
</tr>
<tr>
<td>3. CRITERIA AND MEASUREMENT PROCEDURE</td>
<td>13</td>
</tr>
<tr>
<td>TREATMENT OF DATA</td>
<td>13</td>
</tr>
<tr>
<td>Organization of Data</td>
<td>14</td>
</tr>
<tr>
<td>Validation Procedure</td>
<td>14</td>
</tr>
<tr>
<td>Minimum Optimum Samples</td>
<td>14</td>
</tr>
<tr>
<td>Probability Computations</td>
<td>16</td>
</tr>
<tr>
<td>4. PRESENTATION OF DATA</td>
<td>17</td>
</tr>
<tr>
<td>Validation</td>
<td>17</td>
</tr>
<tr>
<td>Sample Size</td>
<td>17</td>
</tr>
<tr>
<td>Words per T-Unit, Figures 1-5.</td>
<td>20</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Subordinate Clauses per T-Unit, Figures 6-10</td>
<td>26</td>
</tr>
<tr>
<td>Clauses per T-Unit, Figures 11-15</td>
<td>32</td>
</tr>
<tr>
<td>T-Units per Sentence, Figures 16-20</td>
<td>38</td>
</tr>
<tr>
<td>Summary of Presentation of Data</td>
<td>44</td>
</tr>
<tr>
<td>5. DISCUSSION AND IMPLICATIONS.</td>
<td>45</td>
</tr>
<tr>
<td>Validity</td>
<td>45</td>
</tr>
<tr>
<td>The Minimum Optimum Sample Size.</td>
<td>45</td>
</tr>
<tr>
<td>Limitations of Application</td>
<td>46</td>
</tr>
<tr>
<td>Summary</td>
<td>47</td>
</tr>
<tr>
<td>Conclusion</td>
<td>47</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>49</td>
</tr>
<tr>
<td>VITA</td>
<td>53</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table                                                                 Page

    I. Means for the Four Criteria and Grades       18
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Grade 7, Compared to T-Unit Series Number and Student Frequency.</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Grade 10, Compared to T-Unit Series Number and Student Frequency.</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Grade 12, Compared to T-Unit Series Number and Student Frequency.</td>
<td>23</td>
</tr>
<tr>
<td>3</td>
<td>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Grade 14, Compared to T-Unit Series Number and Student Frequency.</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Overall, Compared to T-Unit Series Number and Student Frequency.</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Grade 7, Compared to T-Unit Series Number and Student Frequency.</td>
<td>27</td>
</tr>
<tr>
<td>6</td>
<td>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Grade 10, Compared to T-Unit Series Number and Student Frequency.</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Grade 12, Compared to T-Unit Series Number and Student Frequency.</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Overall, Compared to T-Unit Series Number and Student Frequency.</td>
<td></td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>9. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Grade 14, Compared to T-Unit Series Number and Student Frequency</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>10. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Overall, Compared to T-Unit Series Number and Student Frequency</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>11. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Grade 7, Compared to T-Unit Series Number and Student Frequency</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>12. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Grade 10, Compared to T-Unit Series Number and Student Frequency</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>13. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Grade 12, Compared to T-Unit Series Number and Student Frequency</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>14. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Grade 14, Compared to T-Unit Series Number and Student Frequency</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>15. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Overall, Compared to T-Unit Series Number and Student Frequency</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>16. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Grade 7, Compared to Sentence Series Number and Student Frequency</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>17. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Grade 10, Compared to Sentence Series Number and Student Frequency</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>
18. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Grade 12, Compared to Sentence Series Number and Student Frequency ........................................ 41

19. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Grade 14, Compared to Sentence Series Number and Student Frequency ........................................ 42

20. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Overall, Compared to Sentence Series Number and Student Frequency ........................................ 43
ABSTRACT

The major purpose of this study was to determine how large a sample should be taken in order to reliably estimate student syntactic ability by using T-unit related criteria. Compositions were collected from twenty-nine students in the seventh grade and from thirty students in the tenth and twelfth grades and from thirty college upperclassmen. The mean number of words per T-unit, subordinate clauses per T-unit, clauses per T-unit, and T-units per sentence was computed for each grade and for all grades combined. The mean number of words per T-unit for the seventh grade was correlated with the mean number of words of the first T-units of all the seventh grade compositions. The number of words per T-unit for the seventh grade was next correlated with the mean of two means: the mean number of words for the first and second T-units from all seventh grade compositions. The mean of three means was used for a third correlation. The correlation procedure was repeated to encompass ever-increasing sample sizes from the grade's compositions. Correlation coefficients were recorded and plotted on a graph. The correlation procedure was applied in the same manner for the four T-unit criteria to grades seven, ten, twelve, and to the college students as well as a combination of the four grades.

The results of the study showed that the small samples correlated in the .80's and .90's with the mean of the whole at twenty T-units or twenty sentences, indicating that a reliable estimation of a group's
skill regarding T-unit length, use of clauses, and T-units per sentence can be made by using a mean of the first twenty T-units or the first twenty sentences from all the compositions in the group.
Chapter 1

INTRODUCTION

A concern in education is the measurement of composition ability. Fluent English writing ability has been a goal in American public education for many years. English writing skills were emphasized in Benjamin Franklin's academy in 1751 (Richey, 1973:445). Composition ability has often been measured by subjective, unreliable means (Smith and Adams, 1972:20-22). Concern has been voiced recently pertaining to reliability in composition measurement (Curriculum Report, June, 1976). Recent research has established, however, objective procedures for composition evaluation which, although limited to certain criteria, do lend objectivity to the evaluation of writing skills.

PURPOSE OF THE STUDY

The major purpose of this study was to supplement recent composition research by determining the minimum optimum written sample required to evaluate syntax at differing grade levels according to certain selected criteria. A secondary purpose was to incorporate a method of sample size estimation into composition measurement.

NEED FOR THE STUDY

The literature reviewed in the area of writing and composition evaluation indicated a paucity of statistical procedures for the
estimation of reliable sample size. Although sample parameters of 100 words had been established since the 1940's (Chotlos, 1944), these parameters were too large to permit rapid evaluation. Other studies reviewed neglected the question of sampling (Bear, 1939; Stalnaker, 1934; Veal, 1966; Coffman, 1966; Akeju, 1972; Klein and Hart, 1968). A need for this study was justified because recently developed methods for the objective evaluation of syntactic maturity (Hunt, 1965) require adequate sampling which, in the evaluation of original compositions, has been previously confined to very long compositions.

THEORETICAL BASIS

The theoretical basis of the study was that a statistical technique existed to determine what specific fraction of a written composition adequately represents the syntactic maturity of the whole.

SCOPE OF THE STUDY

The participants came from the seventh, tenth, and twelfth grades at University High School, Baton Rouge, Louisiana, and from college upperclassmen at Louisiana State University. Each of the four groups consisted of thirty members except for the seventh grade, which contained twenty-nine participants.
DEFINITIONS

1. internal consistency: the property of a set of scores or numbers to be free of chance deviations beyond a certain degree. In this study quantitative T-unit data within compositions were tested for internal consistency to produce a small, reliable sample that could be taken from a longer composition.

2. minimum optimum sample size: the smallest and most reliable sample that can be taken from written compositions. The minimum optimum sample is consistent with the group from which it was taken to a degree indicated by a coefficient of correlation.

3. progressive mean: one of a series of arithmetic averages of a set of means that increases N by one with each additional element of the progression.

4. sentence: end punctuation was used in this study to delineate a sentence.

5. syntactic maturity: the ability of a writer measured by the number of words per T-unit, subordinate clauses per T-unit, clauses per T-unit, and T-units per sentence. Research done by Hunt (1965 and 1970) has established that older writers generally produce more words per T-unit, subordinate clauses per T-unit, and clauses per T-unit. Hunt also indicated that T-units per sentence was a fair, but not the best, indicator of syntactic maturity.

6. T-unit: one main clause plus all the subordinate clauses attached to or embedded within it; the minimum terminable unit of the English language.
Chapter 2

REVIEW OF RELATED LITERATURE

The literature reviewed from the 1930's to the present revealed many interesting trends in the field of English written composition measurement. LaBrant, Anderson, Davis, Mann, and Chotlos developed statistical procedures pertinent to this study. Diederich, Veal, Coffman, Akeju, and Klein and Hart developed reliable methods of essay evaluation through the use of grading teams. Slotnick and Page within the last fifteen years have attempted to use computers in grading essays, and others have used objective tests to measure writing ability. Finally, the most recent work, that with the T-unit, shows that a possibility exists for valid measurement of certain criteria. The literature reviewed indicated a need for further research concerning quantitative measurements related to the T-unit.

Among the earliest studies concerning the quantitative measurement of compositions was that of LaBrant (1933). Drawing from general statements by Jean Piaget and Otto Jespersen, LaBrant theorized that use of certain clauses in compositions indicated writing skill. LaBrant's instrument was the subordination index, which was determined by dividing the total number of subordinate clauses by the total number of clauses within a composition. LaBrant found that an increase in subordination index accompanied an increase in chronological age except in groups where chronological age differed greatly from mental age.
Compositions in LaBrant's study averaged 160 words and came from three groups: grades four through nine, grades nine through twelve, and twenty-one eminent psychologists. LaBrant's work is considered by many researchers, such as Stotsky (1976) and Hunt (1965), to be a primary work in the measurement of syntactic maturity.

Frogner (1933) measured the relationship between sentence complexity and level of schooling and found that students composed increasingly more complex sentences as school level increased. Compositions in Frogner's study were generally less than 200 words. Smaller gains were noted between the seventh and ninth grades than between the ninth and eleventh grades.

Anderson (1937) attempted to refine LaBrant's statistical procedures by measuring the consistency of the subordination index within individual compositions. Anderson collected essays written by University of Minnesota English students and intercorrelated the subordination index, the sentence length, and the standard deviation of sentence length for "several" compositions. Anderson was not pleased with his correlation coefficients and concluded that 150 words did not constitute an adequate sample. Anderson demonstrated, however, that correlations of grades assigned by several raters can be quite high. Stalnaker (1934) also found the inter-rater method to be reliable.

Davis (1937), in a companion study to that of Anderson, found that samples of a child's longest sentences are evenly distributed throughout a sample of fifty sentences. Davis selected the five longest sentences from recorded spoken words of children aged five-and-one-half, six-and-one-half, and nine-and-one-half. Davis' definition of a sentence included run-on sentences which reached fifty words in some
instances. Lacking the advantage of the computer, neither Anderson nor Davis could determine the minimum composition or spoken sample required to adequately measure the abilities of the children used in their studies.

Bear (1939) showed that students in grades one through eight write an increasingly greater percentage of complex sentences as they progress through school. Bear reported no data concerning the reliability of samples obtained, but noted that composition length increased with school level. Bear added more evidence to the theory that written verbal ability increases with chronological age, a theory on which modern researchers, notably Hunt, Mellon, and O'Donnel, have constructed their hypotheses.

Fairbanks, Mann, and Chotlos (1944) attempted to develop not only reliable statistical procedures, but also unambiguous evaluative criteria. The researchers jointly developed the type-token ratio (TTR) as an instrument for evaluating compositions. The TTR is the ratio of certain types of words to the total number of words (tokens) in a composition. The three researchers also used improved techniques for sample size estimation.

Fairbanks, in the first of the joint studies, compared the TTR usages of ten psychotic patients from the Mt. Pleasant (Iowa) State Hospital with the TTR usages of ten college freshmen. A sample of 3000 spoken words was recorded for each subject. A mean TTR for each 100 word segment was calculated, giving thirty segments which were then divided at random into two sets of fifteen segments each. The mean TTR of both sets was computed and the $t$ test for differences was applied. The results of the test showed that the internal consistency
of the language samples was, regarding the TTR, significant at the .05 level. Fairbanks arrived at this conclusion because no significant difference existed between the two sets of fifteen.

Mann, using written autobiographies of Fairbanks' group, measured the uses of the traditional eight parts of speech with TTR's. Mann found that no significant difference existed between the combined means of the two groups of words which consisted of randomly selected 1400 word blocks.

Chotlos, in the last of the joint studies, used compositions collected by the Works Progress Administration from a carefully selected sample of Iowa school children. The compositions were a minimum of 3000 words in length. Using Mann's system of dividing the 3000 words into groups of 100, Chotlos used the Pearson r to show that a 1000 word sample was just as consistent, using TTR criteria, as a 3000 word sample. Hunt (1965) relied on a 1000 word sample in his initial study of the T-unit.

Myklebust (1965) also used correlations to determine the adequacy of his sample. He found that when measuring total written errors, the number of errors in the first three sentences correlated significantly with the percentage of errors in the remainder of the paper. Myklebust used criteria that were far broader than any of those cited previously in this review.

All of the above researchers realized that they were measuring primarily verbal ability. Huddleston (1954) criticized other researchers who thought that verbal ability constituted the sole skill in composition. Huddleston noted that the Educational Testing Service was planning to address the fundamental nature of writing ability, a
problem with which ETS psychometrists later became discouraged (Curriculum Report, June, 1976).

Several attempts to define and measure writing ability since the early 1950's have been made, most of which have neglected sample size. One method correlates the ratings of several graders who use the same standards. Stalnaker (1934) was an early developer of this method. Diederich (1974), under the sponsorship of the National Council of Teachers of English, published a booklet describing the inter-rater method. Diederich had worked with the inter-rater method with the Educational Testing Service. The inter-rater correlation method has attained ratios above .90 and has been used successfully with as many as five raters. Follman and Anderson (1967), however, criticized the inter-rater method as having possible drawbacks when homogeneous raters, such as all the students in an English methods class, are used. Follman and Anderson implied that the criteria used by the graders may be uniform but not valid.

Veal (1966) attempted to establish validity by using several raters and comparing student essays with model essays graded by "experts." The inter-rater reliability of the experts ranged from .50 to .63. Inter-rater correlations for Veal's study ranged from .47 to .80.

Coffman (1966) compared the results of inter-rater gradings of essays with the results of an objective composition test. Correlations between four sets of essays and the objective tests were .703, .709, .725, and .710. He was attempting to establish validity for his essay criteria. Coffman carefully stated that if essay tests are to be compared with objective tests, both must correlate with a common reliable
measure. Coffman meant, in essence, that the criterion to be measured must be identified before statements about validity can be made. Stalnaker's (1934) study is an example of research done with loosely defined criteria. Using intelligence test scores and English grade point averages as criteria, Stalnaker concluded that the objective tests used in his study were more valid than the essay tests. "Good writing" was not defined in the study, so the conclusion regarding validity is ambiguous.

Akeju (1972), in his study of students learning English as a foreign language, noted that inter-rater reliability was not satisfactory (.72). He called for the development of a multiple choice test to measure writing ability.

Fagan (1975), after a study of twelve writing-ability tests published since 1960, reported that only three gave reliability data and those three used the inter-rater method.

Klein and Hart (1968) found that grades assigned by law professors on student compositions were quite consistent and the professors showed inter-rater correlations that were quite high. Klein and Hart had difficulty in ascertaining the criteria used by the professors, but they said that the professors apparently preferred longer answers and persuasive, "common sense" answers.

The inference made from the studies reviewed on inter-rater correlation is that high reliability may be obtained in evaluating compositions if several raters are used who have a common criteria with which to evaluate compositions. If the inter-rater method is used, the raters must be practiced in their task. If the essay is compared to an objective evaluation, then the evaluator must be aware of the criteria
used in the evaluation of both the essay and the objective tests.

Slotnick (1972) recognized that not all experts agreed on the criteria required to distinguish good writing from poor writing. Building on the work done by Page (1966), Slotnick attempted to evaluate compositions using a computer. Using a special program, the computer registered thirty-four criteria that were thought to be important factors used by experts in essay grading. The criteria were drawn from various studies and from prescriptive grammar guides. Actual essays done by 476 high school students were fed into the computer. The computer measured the frequency of punctuation marks, the mean sentence length, the mean word length, and the thirty-one other criteria that the study required. Since the computer was acting as a surrogate expert, its criteria, too, could be questioned.

The T-unit, developed by Hunt (1965), represents an attempt to measure syntactic maturity. Building on the work of Bear, Davis, Frogner, LaBrant, and others, Hunt theorized that skill in the use of certain syntactic structures would increase with grade level. After an extensive study of existing methods of evaluation of written compositions, Hunt found the criteria he investigated to be inadequate. He developed a criterion that could measure subordination in sentences without regard to punctuation marks. His original study in 1965 and his related study in 1970 showed that subordination skill as measured by T-unit criteria increased significantly with the grade level of the student. By utilizing the fact that skill increased with grade in school, Hunt established that the T-unit was a valid instrument in evaluating the level of maturity of syntactic structures in a composition. Compositions from the fourth, eighth, and twelfth grades were
used in Hunt's original study.

Loban (1966), working independently from Hunt, developed the "communication unit," which is similar to the T-unit. Loban worked with oral English whereas Hunt used written English. Hunt (1970) acknowledged Loban's contribution.

Mellon (1967) and Bateman and Zidonis (1966) did studies similar to Hunt's using words per T-unit, mean number of clauses per T-unit, and mean number of T-units per sentence as part of their criteria.

Rubin and Buim (1974) did a study of the interrelationship of five parameters of composition evaluation. They found that T-unit length in words was independent of correct usage, thus affirming Hunt's statement that the T-unit was a measure of syntactic maturity and not of usage.

Bivens and Edwards (1974) used the T-unit to show writing improvement as did O'Donnel (1968), who recognized the need for a more accurate method of computing reliability than existed at that time.

The National Assessment of Educational Progress (1975) cited Myklebust in the bibliography of its report. No mention was made as to how Myklebust's statistical procedure for calculating sample size for his general criteria of syntactic quotient was adapted to the NAEP procedure of selecting specific errors and classifying them separately.

In summary, several techniques exist for the measurement of written compositions. Earlier studies concentrated on developing statistical procedures to measure writing ability. Later studies were made to determine the validity of criteria used in evaluating
compositions. The development of the T-unit gave researchers an unambiguous tool that has been verified as an indicator of syntactic sophistication. Some question still exists as to the minimum optimum sample size that can be used in T-unit evaluations.
Chapter 3

CRITERIA AND MEASUREMENT PROCEDURE

The determining of a group minimum optimum sample size using T-unit related criteria was the primary purpose of this study. Students from each of the four grades, thirty from each except for the seventh grade, which had twenty-nine, were shown a projected image of a boy playing and were given an hour to write an essay. The participants were instructed to write an original, anonymous composition using the picture to suggest a theme. All of the compositions were written during the month of September, 1976.

Four criteria were used in evaluating compositions: number of words per T-unit, number of subordinate clauses per T-unit, total number of clauses per T-unit, and the number of T-units per punctuated sentence. A correlation procedure was used to establish the minimum composition length needed to reliably evaluate the compositions of groups at four levels: seventh grade, tenth grade, twelfth grade, and college upperclassmen (coded grade fourteen). The sample estimation was made for groups instead of for individuals because the T-unit studies reviewed had worked with groups only.

TREATMENT OF DATA

The procedure for treating the data was selected after weighing a number of factors. The procedure had to be efficient to use in terms
of time and adequacy of results, and it had to be compatible with the compositions written by the students. A correlation procedure was selected whereby the mean of a small sample was correlated with the mean of the whole.

**Organization of the Data**

T-units and punctuated sentences were identified and numbered in a consecutive series within each student's composition. The series numbering began with the first T-unit or sentence of each composition. Sentences were numbered separately from T-units.

**Validation Procedure**

A partial replication of the original T-unit work done by Hunt (1965) was performed to test the syntactic similarity of the students in the present study with the students in the previously cited studies. The mean number of words per T-unit, mean number of subordinate clauses per T-unit, mean number of clauses per T-unit, and the mean number of T-units per sentence were calculated for each of the four participating grades. The trend shown by the means of the different grades was compared to the trends of the earlier studies. The means derived for the validation procedure also served as the group means in the sample size estimations.

**Minimum Optimum Samples**

Minimum optimum sample sizes were obtained for each of the four criteria categories for the entire 119 students and for each grade, resulting in twenty sample-size estimations.

The internal consistency of the compositions of each of the twenty
estimation groups was computed to find the minimum optimum sample sizes. Two sets of derived data were used in a correlation which yielded coefficients that were adequate for sample size estimation. One set, the group mean, remained constant throughout a single sample estimation computation. Twenty group means were used in the entire study.

The second data set utilized in the correlation was a series of consecutive, progressive means, derived in the following manner:

\[
\begin{align*}
\text{First Mean} & : PM_1 = \frac{X_{m1}}{1} \\
\text{Second Mean} & : PM_2 = \frac{X_{m1} + X_{m2}}{1 + 1} \\
\text{Third Mean} & : PM_3 = \frac{X_{m1} + X_{m2} + X_{m3}}{1 + 1 + 1}
\end{align*}
\]

or, generally,

\[
PM_n = \frac{X_{m1} + X_{m2} + X_{m3} + \ldots + X_{mn}}{N}
\]

and

\[
PM_n = \frac{\sum_{m} X_{mn}}{N}
\]

where \(PM\) is the progressive mean, \(N\) and \(n\) represent the number of means used, and \(X_m\) is the mean of the entire set of T-units or sentences that have a series number corresponding to the subscript number.

The correlation coefficients derived between the group mean, which included a larger number of T-units or sentences, and the progressive means, which included a smaller number of T-units or sentences, showed the relationship between the value of the large sample and the value of the smaller sample. The highest correlation coefficients were judged, within certain limitations, to be the better indicators of
minimum optimum sample sizes. An interpretation of minimum optimum sample size was presented in the form of number of T-units and number of sentences which contributed to the PM at the optimum level.

**Probability Computations**

As the PM progressed through the series of consecutive T-units or sentences, the number of student contributors diminished because some students wrote fewer T-units or sentences than did others. The number of students contributing at each series number was recorded and a cumulative student frequency distribution per series number was established.

The probability of chance occurrence of a greater correlation was computed to facilitate the recognition of the effect of varying numbers of student contributors. The .01 level of probability was arbitrarily set as the level below which a correlation was deemed untenable.
Chapter 4

PRESENTATION OF DATA

The statistical analysis of the data yielded a number of results. The combined group of 119 students produced 11,228 T-units and sentences. Individual means for each of 119 compositions, group means for each of the four grades, and enough progressive means to correlate with the group means 768 times were produced from the data. Probability was also computed from the data.

Validation

A successive increase by grade in the mean length of the first three criteria was compared with the trends shown by Hunt (1965 and 1970). The purpose of the comparison was to partially replicate Hunt's research to determine if the data in the present study were comparable to Hunt's data. An inspection of the data in Table I indicates a hierarchal order of categorical means by grade level in the criteria of words per T-unit, subordinate clauses per T-unit, and clauses per T-unit. The fourth criterion, T-units per sentence, showed irregularities possibly caused by the greater use of run-on sentences by seventh graders and the use of compound sentences by the twelfth graders. The results shown in Table I closely concurred with those of Hunt (1965).

Sample Size

The sample size estimation procedure yielded three distinct and
Table I
Means for the Four Criteria
and Grades

<table>
<thead>
<tr>
<th>Grade</th>
<th>Mean Words/TU</th>
<th>Mean Subordinate Clauses/TU</th>
<th>Mean Clauses/TU</th>
<th>Mean TU/Sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>9.373</td>
<td>.162</td>
<td>1.212</td>
<td>1.168</td>
</tr>
<tr>
<td>10</td>
<td>11.635</td>
<td>.277</td>
<td>1.339</td>
<td>1.037</td>
</tr>
<tr>
<td>12</td>
<td>12.712</td>
<td>.356</td>
<td>1.359</td>
<td>1.159</td>
</tr>
<tr>
<td>14</td>
<td>15.049</td>
<td>.365</td>
<td>1.387</td>
<td>1.014</td>
</tr>
</tbody>
</table>

Interdependent results which are best presented together to ensure clarity.

Figures 1-20 give data on the correlations of progressive means to the group means, on probability, and on the number of students contributing to the correlations at each point. A brief explanation of an example will demonstrate the structure of the graphs.

Figure 1, which is constructed in the same format as the other figures, shows correlation coefficients in the vertical axis, labeled "r." The horizontal axis labeled "A" shows not the progressive means themselves, but the series number of each progressive mean. By selecting a series number, such as the fifteenth, it can be seen that the fifteenth progressive mean correlated with the mean of the whole group to an extent indicated by .90 on the vertical axis. Since the fifteenth progressive mean represents the first fifteen T-units of, in the case of Figure 1, most of the seventh grade, the graph shows that
the mean words per T-unit for the first fifteen T-units correlate with
the mean words per T-unit for all the T-units produced by the seventh
grade to a great degree. "Most" of the seventh grade means, for the
fifteenth T-unit in Figure 1, twenty-two students. Scale B along the
horizontal axis indicates how many students contributed to the corre­
lation. Note can be taken that the entire class, twenty-nine, contrib­
uted T-units from the second T-unit series number to the tenth, where
one student ceased writing. This event is recorded as twenty-eight
directly under the number ten in scale A. At the thirteenth T-unit
three more students ceased writing, an event signified by the number
twenty-five beneath the position for the number thirteen in scale A.
Scale B shows how the number of students contributing to the progres­
sive means and the correlations varied. Scale B does not have any
coordinates on the graph.

Probability bands, shown by the rows of dots in the figures,
indicate where chances are less than or equal to one in one hundred
that a greater correlation than the one shown could have occurred by
chance. An example from Figure 1 will illustrate the function of the
probability bands for all figures. It can be seen that the fifteenth
series number lies under the band of probability. Chances are very
slight that a higher correlation than .90 could have appeared by
accident. In only one case out of a hundred should the correlations
actually shown on the graph be erroneously high. The probability band
in Figure 1 casts doubt on the high correlations obtained after the
thirty-first progressive mean and also on the low correlations up to
the fourth progressive mean.
Figures 1-20 are arranged by criterion category in the following order: words per T-unit, subordinate clauses per T-unit, clauses per T-unit, and T-units per sentence. Each of the four categories presents results from each of the four grades for overall.

Words per T-Unit, Figures 1-5

Figures 1-5 illustrate the positive correlations of the progressive means with the mean of the whole group for the category of mean words per T-unit. Grades seven, ten, twelve, fourteen and overall are represented. Figures 1-5 indicate that an adequate correlation is reached after the first ten T-units and a very good correlation is reached after the first twenty.

Figure 5, which is an aggregate figure, is smoother than the first four figures because of the effect of the larger group.
Figure 1. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Grade 7, Compared to T-Unit Series Number and Student Frequency
Figure 2. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Grade 10, Compared to T-Unit Series Number and Student Frequency.
Figure 3. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Grade 12, Compared to T-Unit Series Number and Student Frequency

Legend

----- Probability < .01 of r>r shown

A T-Unit Series Number
B Cumulative Student Frequency Distribution per Series Number (in Decreasing Order)
Figure 4. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Grade 14, Compared to T-Unit Series Number and Student Frequency
Figure 5. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Words per T-Unit, Overall, Compared to T-Unit Series Number and Student Frequency
Subordinate Clauses per T-Unit, Figures 6-10

Although the curves in the category of subordinate clauses per T-unit are less smooth than those for words per T-unit, the subordinate clause curves also indicate that a reliable sample may be taken at around twenty T-units.

Figure 10 illustrates the averaging effect of the large group as did Figure 5.
Figure 6. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Grade 7, Compared to T-Unit Series Number and Student Frequency
Figure 7. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Grade 10, Compared to T-Unit Series Number and Student Frequency
Figure 8. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Grade 12, Compared to T-Unit Series Number and Student Frequency.
Figure 9. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Grade 14, Compared to T-Unit Series Number and Student Frequency
Figure 10. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Subordinate Clauses per T-Unit, Overall, Compared to T-Unit Series Number and Student Frequency
Clauses per T-Unit, Figures 11-15

Clauses per T-unit included all relative clauses occurring within the T-unit plus the subordinate clauses. A comparison of each grade's curve in this category with the same grade's curve in the category of subordinate clauses per T-unit revealed some similarities.

The criterion of clauses per T-unit yielded reliable samples at the same approximate levels as the previously illustrated criteria: twenty T-units.
Figure 11. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Grade 7, Compared to T-Unit Series Number and Student Frequency
Figure 12. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Grade 10, Compared to T-Unit Series Number and Student Frequency
Figure 13. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Grade 12, Compared to T-Unit Series Number and Student Frequency
Figure 14. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Grade 14, Compared to T-Unit Series Number and Student Frequency.
Figure 15. Positive Correlations of Consecutive, Progressive Means of Serially Numbered T-Units with the Group Mean for the Category of Number of Clauses per T-Unit, Overall, Compared to T-Unit Series Number and Student Frequency
T-Units per Sentence, Figures 16-20

The last category in the study, T-units per sentence, differed from the other categories in that it was a matter of punctuation rather than a measure of ability to construct complex T-units. The category of T-units per sentence served as a demonstration of the consistency of reliable sample size throughout this study: approximately twenty sentences constituted a fairly reliable sample.

The category of T-units per sentence, being different from the other categories, gave rise to a question concerning the structural consistency of the English language: how many other language units can be sampled with the means of about twenty samples?
Figure 16. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Grade 7, Compared to Sentence Series Number and Student Frequency
Figure 17. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Grade 10, Compared to Sentence Series Number and Student Frequency
A Sentence Series Number
B Cumulative Student Frequency Distribution per Series Number (in Decreasing Order)

Figure 18. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Grade 12, Compared to Sentence Series Number and Student Frequency
Figure 19. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Grade 14, Compared to Sentence Series Number and Student Frequency
Figure 20. Positive Correlations of Consecutive, Progressive Means of Serially Numbered Sentences with the Group Mean for the Category of Number of T-Units per Sentence, Overall, Compared to Sentence Series Number and Student Frequency
Summary of Presentation of Data

The data presented in the present chapter lend themselves to a study of sample size using T-unit data. Table I illustrated data that aided in the validation of the study. The twenty figures showed a correlation of means from specific samples with the group mean. The figures also presented the number of students contributing to each correlation and the possibility of a greater correlation occurring by chance.
Chapter 5

DISCUSSION AND IMPLICATIONS

The central question, that of finding the minimum optimum sample size for groups was, within the limits of this study, answered. The data presented in the previous chapter also contributed to the validation of the study and showed a number of concepts concerning sample size which apply to situations beyond this study.

Validity

A sufficient replication of the work of Hunt and others was demonstrated by the successive increase in the means of the first three T-unit criteria throughout all grade levels.

The fourth criterion, T-units per sentence, probably was erratic for several reasons. The seventh grade mean probably exceeded the means of the other grades because the seventh graders tended to use more run-on sentences. The twelfth grade probably had a higher mean than the foruteenth grade because the twelfth graders used more compound sentences than did the fourteenth. Further study is needed to verify the reasons for the differences.

The Minimum Optimum Sample Size

A survey of the data indicates that an optimum sample for a group's production of T-unit data can be obtained within a comparatively wide band of T-unit quantities. The data show that, regarding
the samples taken in this study, the T-unit and sentence sample sizes
with a probability of \( \leq .01 \) of not having a greater chance correlation are indicators of optimum sample sizes at their respective
correlation levels as shown in Figures 1-20. The selection of an
exact quantity to T-units or sentences to use for a minimum optimum sample size is determined largely by the correlation coefficient
desired. All of the graphs shown give a point where the curve begins
to rise less steeply. The points indicate places where minimum optimum samples can be taken. These crucial bending points lie in the
.80's and lower .90's, providing a fairly high coefficient of correlation. Of course, optimum samples appear well after the points marking
the beginning of the small rise, but samples nearing correlations such as .95 are often larger and would produce diminishing returns for the amount of work required to take them. In general the results of this study show that the criteria considered herein may safely be sampled from groups of thirty or more by collecting data from the first twenty T-units or sentences and computing a mean. Within the limits of this study, a sample taken in this manner is representative of the whole.

Limitations of Application

The T-unit studies reviewed used groups instead of individuals. The original purpose of using T-units was to measure the syntactic maturity at different grade levels and general statistical considerations dictated that groups of students be used instead of only one student. This study has also been limited to groups, and its results should not be applied to the sampling of individual compositions.
Further study should be done concerning the sampling of individuals.

The study was also limited by the very nature of the T-unit. T-units imply syntactic sophistication to some degree in writing. Subjective and logical evaluations are not tested by T-units.

Summary

The task of estimating minimum optimum sample size in the compositions of groups was recognized as being important to measuring the syntactic sophistication of large groups with T-unit criteria. T-units were defined as minimum terminable subject-verb structures. Minimum optimum sample size was estimated for four criteria: words per T-unit, subordinate clauses per T-unit, clauses per T-unit, and T-units per punctuated sentence. College upperclassmen, seventh, tenth, and twelfth graders wrote compositions for the study.

A mean for each category was computed at each grade level and correlated with a progressive mean a number of times. The progressive mean was a constantly changing figure that began with the means of all the first T-units or sentences in a group for the criterion to be considered. The next number in the progressive mean was a similarly computed mean from the second T-units or sentences from the group averaged with the mean of the first T-units or sentences. The progression continued in the same manner until all T-units or sentences were examined from the group.

Conclusion

According to the criteria and techniques used in this study, the means of the first twenty T-units or sentences from each composition
in a group of thirty or more are, for practical purposes, a satisfactory sample of the compositions of the entire group.
LITERATURE CITED

1. Books


2. Periodicals


Veal, L. Ramon. "Measuring Writing Improvement During an NDEA English Institute," *Journal of Educational Measurement*, 3, No. 4, (Winter,


VITA

James Howard Wynn grew up in North Louisiana. He has taught junior high school and has worked as a graduate assistant at Louisiana State University. He is married to the former Gloria Ann Higgins.
EXAMINATION AND THESIS REPORT

Candidate: James Howard Wynn

Major Field: Education

Title of Thesis: DETERMINING THE INTERNAL CONSISTENCY OF ENGLISH COMPOSITIONS USING SELECTED CRITERIA

Approved:

[Signatures]

Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

[Signatures]

Date of Examination:

March 24, 1977