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.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106
FADALA, Sam Nicholas, 1939-
DEVELOPMENT OF COMPOSITION SKILLS IN
THE UNIVERSITY OF ARIZONA MODEL FRESHMAN
COMPOSITION PROGRAM.

The University of Arizona, Ed.D., 1974
Education, secondary

Xerox University Microfilms, Ann Arbor, Michigan 48106

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.
DEVELOPMENT OF COMPOSITION SKILLS IN
THE UNIVERSITY OF ARIZONA MODEL
FRESHMAN COMPOSITION PROGRAM

by
Sam Nicholas Fadala

A Dissertation Submitted to the Faculty of the
DEPARTMENT OF SECONDARY EDUCATION
In Partial Fulfillment of the Requirements
For the Degree of
DOCTOR OF EDUCATION
In the Graduate College
THE UNIVERSITY OF ARIZONA

1974
I hereby recommend that this dissertation prepared under my direction by Sam Nicholas Fadala, entitled Development of Composition Skills in The University of Arizona Model Freshman Composition Program, be accepted as fulfilling the dissertation requirement of the degree of Doctor of Education.

Robert J. Lemon
Dissertation Director
Date: July 11, 1974

After inspection of the final copy of the dissertation, the following members of the Final Examination Committee concur in its approval and recommend its acceptance:

Richard A. King

Fred W. Cell

Robert J. Lemon

James R. Rankin

Date

2/11/74

7/11/74

7/11/74

7/11/74

This approval and acceptance is contingent on the candidate's adequate performance and defense of this dissertation at the final oral examination. The inclusion of this sheet bound into the library copy of the dissertation is evidence of satisfactory performance at the final examination.
STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: [Signature]

[Name]
ACKNOWLEDGMENTS

I wish to express my gratitude to the following for their unselfish support: Dr. Robert Letson, my dissertation director, Dr. Raymond Klein, Dr. James Rankin, Dr. Richard Krebs, and Dr. Robert Grant, members of my committee, Mr. and Mrs. Nicholas Fadala, my parents, and Nancy Fadala, my wife. I wish to thank also Henrietta Maykulsky for her patient aid, and Mrs. Robert Letson for her interest in my work.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>I INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>2</td>
</tr>
<tr>
<td>Significance of the Problem</td>
<td>3</td>
</tr>
<tr>
<td>Assumptions</td>
<td>4</td>
</tr>
<tr>
<td>Hypotheses Tested</td>
<td>6</td>
</tr>
<tr>
<td>Definitions of Terms</td>
<td>6</td>
</tr>
<tr>
<td>Summary</td>
<td>8</td>
</tr>
<tr>
<td>II REVIEW OF THE LITERATURE</td>
<td>9</td>
</tr>
<tr>
<td>Topic Selection</td>
<td>9</td>
</tr>
<tr>
<td>Rating Scales in the Evaluation of Composition Skills</td>
<td>13</td>
</tr>
<tr>
<td>The Employment of Graders and Their Reliability in Evaluating Written Composition</td>
<td>18</td>
</tr>
<tr>
<td>The Use of Criteria as Delineators of Composition Evaluation</td>
<td>22</td>
</tr>
<tr>
<td>Review of the Specific Criteria Selected for the Study</td>
<td>29</td>
</tr>
<tr>
<td>Organization</td>
<td>35</td>
</tr>
<tr>
<td>Mechanics</td>
<td>37</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>40</td>
</tr>
<tr>
<td>Communication of Idea</td>
<td>41</td>
</tr>
<tr>
<td>Logical Developmental Sequence</td>
<td>43</td>
</tr>
<tr>
<td>Summary</td>
<td>45</td>
</tr>
<tr>
<td>III PROCEDURES</td>
<td>47</td>
</tr>
<tr>
<td>The Research Design</td>
<td>47</td>
</tr>
<tr>
<td>The Population</td>
<td>48</td>
</tr>
<tr>
<td>Assurance of Student Anonymity</td>
<td>49</td>
</tr>
<tr>
<td>Collection of Student Data</td>
<td>50</td>
</tr>
<tr>
<td>Criteria Selection</td>
<td>51</td>
</tr>
<tr>
<td>The Test Instrument</td>
<td>51</td>
</tr>
<tr>
<td>Selection of Graders</td>
<td>52</td>
</tr>
<tr>
<td>TABLE OF CONTENTS--Continued</td>
<td></td>
</tr>
<tr>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>Establishment of Reader Reliability Coefficient Using the Pearson Product Moment Correlational Technique</td>
<td>52</td>
</tr>
<tr>
<td>Topic Selection</td>
<td>54</td>
</tr>
<tr>
<td>Process for Administering Pretest and Posttest Themes</td>
<td>55</td>
</tr>
<tr>
<td>Process for Scoring the Pretest, Posttest Essays</td>
<td>55</td>
</tr>
<tr>
<td>Analysis of the Data</td>
<td>56</td>
</tr>
<tr>
<td>Summary</td>
<td>57</td>
</tr>
<tr>
<td>IV FINDINGS</td>
<td>60</td>
</tr>
<tr>
<td>Percentage of Increase on Each Individual Study Criterion in Terms of Accumulated Points</td>
<td>61</td>
</tr>
<tr>
<td>Results of the Correlated t-test of Two Means for Establishing Statistical Level of Significance for Pretest and Posttest</td>
<td>63</td>
</tr>
<tr>
<td>Results of the Correlated t-test of Differences Between Two Means Based on the Eight Student Variables</td>
<td>63</td>
</tr>
<tr>
<td>Results of the Step-Wise Multiple Linear Regression Analysis of the Variance on the Posttest Dependent Variable</td>
<td>72</td>
</tr>
<tr>
<td>Reader Reliability Levels for the Pretest and Posttest</td>
<td>74</td>
</tr>
<tr>
<td>Summary</td>
<td>78</td>
</tr>
<tr>
<td>V SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS</td>
<td>83</td>
</tr>
<tr>
<td>Summary</td>
<td>83</td>
</tr>
<tr>
<td>Conclusions</td>
<td>88</td>
</tr>
<tr>
<td>Recommendations for Future Applications and Research</td>
<td>89</td>
</tr>
<tr>
<td>APPENDIX A: STUDENT INFORMATION QUESTIONNAIRE</td>
<td>91</td>
</tr>
<tr>
<td>APPENDIX B: EVALUATION SHEET</td>
<td>92</td>
</tr>
<tr>
<td>APPENDIX C: CRITERIA DEFINITIONS</td>
<td>93</td>
</tr>
<tr>
<td>APPENDIX D: TOPIC SELECTION QUESTIONNAIRE</td>
<td>95</td>
</tr>
<tr>
<td>APPENDIX E: PRETEST ASSIGNMENT $O_1$</td>
<td>96</td>
</tr>
<tr>
<td>---------------------------------------------------------</td>
<td>----</td>
</tr>
<tr>
<td>APPENDIX F: POSTTEST ASSIGNMENT $O_2$</td>
<td>97</td>
</tr>
<tr>
<td>LIST OF REFERENCES</td>
<td>98</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table

1. Pearson Reader Reliability Correlation Coefficient for Pre-Study Sample Themes ........................................... 53
2. Grader-Awarded Raw Point Percentage Gains Per Individual Study Criterion ................................. 62
3. Correlated t-test of Two Means for the Pretest-Posttest ................................................................. 64
4. Delineation of Student Data Variables for the Correlated t-test ....................................................... 66
5. Correlated t-test of Two Means on the Eight Student Variables ....................................................... 67
6. Results of the Step-Wise Multiple Linear Regression Analysis with the Posttest as Dependent Variable .......... 73
7. Delineation of Student Data Variables for the Step-Wise Linear Regression Technique .......................... 75
8. Pearson Reader Reliability Correlation Coefficient for Pretest Themes ............................................. 76
9. Pearson Reader Reliability Correlation Coefficient for Posttest Themes ............................................. 77
10. Variance Accountability on the Posttest Dependent Variable ......................................................... 81
ABSTRACT

This study was undertaken as an investigation of student skill acquisition in The University of Arizona Model Freshman Composition Program.

Major purposes of the study were: (1) to determine if five graders of similar background could recognize changes in $O_1$ and $O_2$ student-composed themes by means of a five-point Likert-type scale based on five criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence, (2) to determine if changes in $O_1$, $O_2$ themes could be assessed in terms of eight student variables: sex, grade point average in high school English, A.C.T. score in English, age, bilingualism--non-bilingualism, class attendance, economic rank, and high school graduation rank, and (3) to effect a model which could serve as a guide for basic evaluation of a given composition program.

The sample included fifty subjects from three classes which had been selected at random from twenty-two sections of Freshman Composition at The University of Arizona.

The design of the study was the one-group, pretest-posttest paradigm, and the data were collected by four means: (1) the administration of pretest and posttest fifty minute in-class writing assignments, (2) a questionnaire designed...
to gather three data: age, language(s) spoken in the home, and family income level, (3) a search of the records on file with The University of Arizona Registrar in order to gather grade point average in high school English, A.C.T. score in English, and graduation rank, and (4) a scanning of the class record books for the purpose of obtaining the sex of the pupil and class attendance record.

The data were analyzed by the following five measures: (1) a simple compilation of raw data to determine percentage increases in Grader-awarded points on each of the five study criteria, (2) a correlated t-test of the difference between two means in order to establish the statistical level of significance between the pretest and the posttest, (3) a correlated t-test of the differences between two means in order to establish the statistical level of significance between the pretest and the posttest concerning each of the eight student variables (above), (4) a step-wise multiple linear regression technique for the purpose of accounting for the variance on the posttest in relation to each of the eight student variables (above) and the pretest, and (5) a Pearson product moment correlation measure to establish reader reliability on the pretest and on the posttest.

The results of the analyses established: (1) an increase in Grader-awarded raw scale points from the 0₁ to 0₂ theme evaluations, (2) a two-tailed probability beyond .001
for the pretest-posttest relationship, (3) a two-tailed probability beyond .029 for the eight student variable changes from the pretest to the posttest, with the exception of age above twenty years, which computed at .290, (4) an assignment of 47.52 percent of the variance on the posttest dependent variable credited to the pretest independent variable, with minute percentages attributed to the eight student variables, and (5) a reader reliability of .73 for the pretest and .43 for the posttest.

The study presented a possible model for composition evaluation.
CHAPTER I

INTRODUCTION

Evaluative research in written composition, because of numerous attending variables, has been considered a generally fruitless endeavor by many scholars in the field. A comprehensive study by an entire staff of this country's most respected minds in the area of composition instruction reviewed a multitude of research efforts and decided that none was sufficiently commendable to deserve notable recognition or emulation.

The scholars were united under the auspices of the National Council of Teachers of English and their task was to criticize the research studies in composition and make intelligent recommendations for future investigations.

A committee of ten began by confronting the entries of some twenty bibliographies of research. More than a thousand citations were reduced to 485 items. Seventeen authorities were consulted to decide which studies should be read. From this consultation—and after the appointment of directors—some 100 titles emerged as worthy of continuing attention. These were then submitted to eighteen other specialists, and 50 additional titles emerged. The three directors of the project now began re-reading the recommended studies. They themselves met as a committee, and during Christmas vacation they consulted the ad hoc committee. Five studies survived the long and learned scrutiny . . . (Hagstrum 1964, p. 54).
Considering the five supposedly best studies, Hagstrum (1964, p. 56), who reviewed the report of the National Council of Teachers of English entitled *Research in Written Composition*, (Braddock, Lloyd-Jones, and Schoer, 1963), agreed with the researchers that "None of these five best studies is definitive. Each raises more problems than it solves."

A plausible reason for lack of definitive results in the research studies is the fact that there has been little agreement among professionals concerning the criteria which underlie acceptable writing and the evaluation of composition programs. As Larson (1970, p. 393) suggested, "... teachers' preferences in writing differ: what one teacher praises and seeks in his students' work may not appear at all among the values of another teacher."

**Statement of the Problem**

The purpose of this study was twofold:

1. The first aspect of this study was to investigate a composition program under specific circumstances in order to determine the possibility of detecting gains in established areas by students who had been compartmentalized into small groups according to predetermined factors.

2. The second aspect was to create a model for emulation, with amendments, for those professionals who
would be involved in teaching composition and evaluating their own programs by employing a set of specific criteria with a rating scale for themes, both of which could be modified to suit the needs of the program under observation.

**Significance of the Problem**

While it has been the contention of some scholars that evaluation of composition programs is at best an unscientific endeavor rendering little data of a conclusive nature, the fact remains that such evaluation must continue, improve, and become an integral part of every writing course due to the universality of the discipline. The writer believes that the most significant form of investigation will prove to be that conducted by the originator of the program in question, and that through experience, intuition, and data uncovered from inquiry, the program can be upgraded to the benefit of each participating student.

An intention of this research was to gain evidence which would serve as a foundation for future studies (specifically in the area of writing instruction referred to as "Freshman Composition") by determining student accomplishment in a particular course. In this case, The University of Arizona Model Freshman Composition Program was studied.
Assumptions

This study, which incorporated The University of Arizona Model Freshman Composition Program, was based on the following assumptions:

1. It was assumed that the second semester students composing the set of subjects for this investigation exhibited sufficient typicality so that application of data from this study to other students in other writing programs would be possible.

2. It was assumed that the design of this investigation would offer a model which could be employed in amended form by those scholars who would desire to duplicate the study, or parts thereof, either with the form utilized by the researcher, or with another paradigm, such as a controlled model perhaps of the test-group, control-group type where such was practical and allowable. In addition to permitting duplication of the study, it was assumed that the basic design could function as a direct model for specific program evaluation.

Limitations of the Study

The following limitations of the study were anticipated:

1. The study was limited to the pupils enrolled at The
University of Arizona in the Freshman Composition Program for the second semester session.

2. The course instructors were limited to those employed by the English Department of The University of Arizona under the direction of Dr. Charles Davis and dedicated to the general prescriptions of The University of Arizona Model Freshman Composition Program.

3. The criteria for evaluation were limited to the five selected by the investigator on the basis of data from the literature, these being: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence.

4. The rating of the five criteria was limited to a five-point scale of Likert design.

5. The evaluation of pretest and posttest themes was limited to the decisions of the five professional persons referred to as "Graders" for this study and selected on the basis of experience in both writing and the teaching of writing. For convenience, the Graders were designated: A, B, C, D, and E.

6. The subjects of the study were compartmentalized into small groups limited to the following criteria: sex, age, grade average in high school English, high
school graduating class rank, bilingualism--non-bilingualism, family income, A.C.T. score in English, and attendance record during the second semester session in The University of Arizona Model Freshman Composition Program.

Hypotheses Tested
The hypotheses which gave direction to this study were:

\[ H_1: \text{There will be no difference between the pretest and posttest themes in terms of the criteria.} \]

\[ H_2: \text{There will be no difference among the small groups of subjects which were compartmentalized on the basis of: sex, age, grade average in high school English, high school graduating class rank, bilingualism--non-bilingualism, family income, A.C.T. score in English, and attendance record in The University of Arizona Model Freshman Composition Program in terms of the pretest and posttest themes.} \]

Definitions of Terms
The following terms were defined within the context of their use in this study:

1. Composition was defined as written expression.

2. The University of Arizona Model Freshman Composition Program was defined as the course in operation
during the time of the study, second semester session, 1974, at The University of Arizona under the direction of Dr. Charles Davis.

3. **Graders** were defined as the five professional persons selected on the grounds of experience in writing and instructing writing, and who were responsible for evaluating the 0₁--0₂ pretest and posttest themes with a five-point scaled instrument (Grading Instrument) and a five criteria format.

4. **Pretest and Posttest Themes** were defined as the fifty minute in-class writing assignments accomplished by the subjects of the study, composed on student selected topics.

5. **Topics** were defined as those writing subjects which were compiled on the basis of a questionnaire given to several classes, including the Freshman level, Senior level, Graduate level, and adult level.

6. **Grading Instrument** was defined as the tool used by the Graders based on a five-point Likert-type scale with five criteria.

7. **Criteria** were defined as the five areas which were rated by the Graders on the Likert type five-point scale, these being: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence, as elaborated in Appendix C.
Summary

A review of hundreds of research studies in the field of composition instruction by a learned group of scholars disclosed that few of the investigations could be considered definitive or worthy of emulation.

However, with some form of composition being taught in every educational institution in the nation, continued, improved, and more widely dispersed research efforts seem imperative.

The purpose of the study was to examine a composition program by employing a foundation of five criteria rated on a Likert-type scale of five levels by professional persons well established in the area of writing instruction.

The study was considered to be significant since the results should prove beneficial to instructors who would survey their own composition programs using the paradigm of this effort as a foundation. Further, there should be value in determining which particular small groups of students gain at what levels on the five criteria of the study: sex, age, grade average in high school English, high school class rank, bilingualism--non-bilingualism, family income, A.C.T. score in English, and attendance record in The University of Arizona Model.
CHAPTER II

REVIEW OF THE LITERATURE

The review of the literature, which is the subject of this chapter, considered five areas. These five areas, presented below, were aspects of the study which were central to the investigation.

1. The effect of theme topic selection on written composition.
2. Rating scales in the evaluation of composition skills.
3. The employment of graders and their reliability in evaluating written composition.
4. The use of criteria as delineators of composition evaluation.
5. Review of the specific criteria selected for the study.

**Topic Selection**

The selection of a theme topic can affect the organization, mechanics, vocabulary, communication, and logical sequence of a theme, as well as other aspects of writing (Cohen 1963, pp. 36-38). Hagstrum (1964, p. 55) believed that "For people learning to write—and even for seasoned writers—the topic determines virtually everything."
While it has been clearly recognized that topic selection could greatly affect the writing product, a particular method of selecting a writing subject has not been decided upon by the scholars involved.

Basically, there have been two schools of thought. The first reasoned that topics should be assigned, the second that topics should generate from the pupil who was to do the writing. The first group often compiled long lists of instructor-designed topics, such as "My Family," "One of My Dreams," "What I Am Afraid Of," and others (Knudsen and Christensen 1957, pp. 101-103). The nomenon of that segment of scholars was to the effect that teacher-assigned subjects would be so constructed as to lead the pupil to commendable responses in his writing efforts, thus promoting the goals of instruction (Farrell 1969, pp. 428-431). Another reason for instructor-prescribed topics had been to promote the melding of literature and composition, for reasons of interrelation, "... to unify English instruction so that the literature and composition are taught together (Kaplan 1969, pp. 1194-1198)." Also, it had been assumed that students would later have to write on subjects not of their choosing, but rather those thrust upon them by future employers or teachers (Larson 1974, p. 80).

Conversely, other instructors of writing, Arntson (1961, p. 4) among them, felt that teacher-forced subjects
did not lend themselves to modern pedagogy and were outdated. The belief was that "Only when you write about something you find interesting can you hope to produce a piece of writing that will be interesting to read (Arntson 1961, p. 4)."

Further, many modern experts had agreed that themes must be written on realistic, meaningful topics in order to display the true ability of the student writer. When the papers were aimed at a problem or condition that realistically involved the lives of the pupils, the themes would be reader-oriented, not grader-oriented, for the students would be attempting to say something in print to a real audience, and not merely to a marker of papers. "For decades," complained Macrorie (1968, p. 686), "we have been smearing bloody marks (sp, awk, gr) in the margins of what we call 'themes.' These papers are not meant to be read, but corrected."

The scholars who called for student-initiated topics insisted that pupils would perform at a higher level of competence when allowed to choose their own subjects (Judy 1970, p. 215). The belief was that "... writing should have some realistic purpose for both the student and the reader. Otherwise the thing remains a purely artificial exercise ..., (Mills 1953, p. 23)," and if artificial, then certainly not in the highest caliber of the student's ability.
A compromise between the two schools of thought on topic selection could be made by allowing students to select topics from a large list which was compiled from student questionnaires, but edited by instructors. The instructor narrowed, defined, and arranged the topics to best serve the purposes of the course. Thus, students would still select subjects which they could honestly expound on, yet which had been modified by the instructor to meet the objectives of the program (Draper 1969, p. 248).

The above compromise would allow students to choose subjects for writing which were within the scope of their experience and ability, and no person would be expected to compose from material not included in his own body of experiences, which was in accord with certain pedagogical principles (Dewey 1938). At the same time, the instructor would take part in the activity, lending to the process his expertise.

Since, as shown by the literature, topic selection was a decisive force behind the motivational attitude of a writer, then certainly it was important and necessary to present writing subjects in a precise and deliberate manner. This aspect of the literature survey was the moving force behind the method in which topics were selected for the $O_1$ and $O_2$ themes of the study, as presented in Chapter III.
Rating Scales in the Evaluation of Composition Skills

While some scholars of the early twentieth century were calling for a continuance of the memorization--recital--rules-of-grammar approach to teaching the mastery of composition skills, insisting that emulation of foreign systems, such as the French, would stand as a panacea for curing the writing problems of American youth (Brown 1915), others were attempting to develop less subjective methods for teaching and evaluating the writing process.

Among these were two scholars, who, working in conjunction, arrived at a method which at least was partially objective in the rating of composition skills. The scholars were M. B. Hillegas and Edward L. Thorndike of Columbia University. The Hillegas-Thorndike scale was an attempt to rank a given number of themes according to merit.

Thorndike (1911, p. 361) considered that "... an ideal scale for merit in English writing would consist of a series of compositions which range from one that possessed zero merit to one that possessed the utmost possible merit...." The purpose of rating themes was, at that time, to grade individual ability, and the aim was to incorporate the use of a true ratio type scale (Minium 1970, p. 18). Positive and negative criticisms of the Hillegas-Thorndike scale were soon to follow its inception.
In 1920 the accuracy of the scale was tested by M. J. Van Wagenen (1920, p. 449) who found that "A group of no more than eight competent judges, working as a unit, can grade at least sixty out of one hundred themes with an error of not more than two points and at least eighty out of the one hundred with an error of not more than three points on a scale of one hundred points."

Without the use of a scale, the same graders using a percentage calculation, arrived at a much lower level of agreement. Thus, Van Wagenen (1920, p. 443) concluded that "With practice in the use of the scale the amounts of error for the same percents would be greatly decreased . . . ."

By 1924 at least one scholar was calling for "homemade composition scales" which could be used by teachers of writing. These scales were to be similar to the Hillegas-Thorndike models in that they considered comparative student writing merit the key, but only seven levels were used, from Merit I to Merit VII (Wilson 1924, pp. 165-170). The student who received a Merit VII rating level would know that he had achieved a high rank compared with his group. Conversely, a Merit level of I would have indicated a woeful lack of competence in writing.

Still, and for many years following, the major purpose of the ranking system was to grade the students in competition with each other. There was only small concern
that the rating method might be advantageous to student
learning, as was casually expressed by Noyes in late 1912
(p. 536). He intimated that "Even the pupil may profit
by studying such a scale, for it will show him just what he
is expected to accomplish and enable him to measure his own
progress."

While Noyes and others did recognize that there
might be some value for the student in ranking his own
writing competency, the thrust of such ranking was still to
categorize the pupil in relation to his peers. Evaluation
of writing was not accomplished with the major aim of a
teaching tool function. However, later in time the purpose
of evaluating student compositions did become partially
instructional, and many scholars later insisted that in-
struction was the only valid reason for such evaluation to
take place at all (Koclanes 1961, pp. 252-264).

In fact, Dusel (1956, p. 327) believed that "... a
meticulous grading of each pupil's efforts against some
standard of excellence may not be desirable." This ex-
pressed, perhaps, the more modern view of the grading
rationale, a diversion from the ranking process for pupil
comparison to a heuristic reason for evaluation. Thus, it
was deemed more valuable to clearly show the student which
areas of composition were in need of improvement, and if
possible, how much improvement was necessary.
Dusel (1956, p. 327) further explained that employing the system of grading compositions on a comparative scale of peers was conducive to possible damage of instruction aims. He said "According to the conventional grading curve, most pupils in an ordinary class are condemned to receive mediocre grades; a sizeable minority will receive D's or F's." The low marks, Dusel believed, would not of themselves further writing competence of the students, but could prove an abasement to them.

Into the present, the majority of instructors of writing have evaluated the products of their students in a subjective manner, and usually with a single summary mark, often a code letter from A to F to designate the level of competency of the paper. There were at least two major problems attending this process.

First, the summary approach, in which the writing was evaluated in toto, led to teaching of the same nature. That is, the subject was presented either in one single unit, or possibly two, generally "mechanics" and "communication." The antithesis to this would have been breaking the discipline down into its many parts and investigating each segment individually.

The second problem of the summary approach was that a sequentially designed teaching plan was denied through that method. Rather than the aspects of the writing process being
broken down into intelligible units for better understanding, everything was considered at one time.

Conversely, by using criteria in conjunction with a rating scale, it could have become a simple process to study each segment of the discipline in detail and in sequence, beginning, perhaps, with that criterion which most disturbed the student, and encouraging his improvement on each individual aspect of the writing process until the entire operation had been enhanced.

The employment of a rating scale would allow the instructor and the student to see which areas of composition were in most need of aid, and to what degree improvement would be required. Rather than a summary mark of A to F, the student might discover that in a specific area his writing was quite competent, while in another it was sadly lacking. Methods of improvement could be concentrated heavily on that area which required the most attention.

Rogal (1966, p. 74) suggested that a major aim of evaluation was to "... fully explain how the paper and the student's general writing ability can be improved." However, it could prove defeating for the student to receive a paper which was intricately marked, often in cryptic fashion, to indicate each and every writing fault in detail.
A rating scale based on specific criteria eliminated this problem, for instead of every aspect of the writing act being attended to, only the major areas of importance were considered, and then, instead of simply a mark to show that something was less than acceptable, the criterion received a numerical score which more precisely showed the student not only that he needed improvement on that criterion, but also how much improvement was necessary according to the judgment of the instructor.

Further, the rating scale instrument, when used with deliberate care in a precise manner based upon preestablished methods, yielded a high reader reliability level, reaching or exceeding .80 when the task was simple rating or ranking (Diederich 1967, p. 577). Thus, the pupil would receive an evaluation of his writing which would more closely agree with a professional consensus, rather than a highly subjective single summary mark of A to F.

The Employment of Graders and Their Reliability in Evaluating Written Composition

Early in this century, scholars recognized that grader evaluation of student themes was at best a haphazard affair with less reliability and agreement than those who instruct prefer. Briggs (1922, p. 423) argued that "... they are surprisingly inconsistent when marking the same
theme a second time after a sufficient interval to insure forgetfulness of the mark first given." However, Briggs (1922, p. 426) later admitted that "... 91 of 100 judges can perceive that theme C is better than theme D ... ,"

in a study of ranked papers.

The apparent disparagement in Briggs' two opposing views of grader reliability pivoted around the evaluation method. Certainly, when allowed a totally subjective approach to the evaluation of student writing, the reliability of a group of graders would be low. When the same graders, however, were given specific criteria to consider, with well established directions for marking, and a rating scale, reliability would rise appreciably.

That graders could disagree widely when placing a mark to indicate value on a theme had been a well established point (Larson 1970, pp. 393-4). This fact led many to believe that the grading of compositions had been, at best, a purely subjective activity centered on the whim of the grader, and that if the same grader marked a theme for a second time after a sufficient lapse of time, he would likely give the theme a different mark as Briggs had indicated.

However, the vast differences in grading consistency, this writer believes, had been due to the fact that graders seldom had parameters established to guide the
process of evaluation. The literature disclosed that when graders did have a clearly defined route to follow, providing that they were of similar professional background, there would be a degree of reliability sufficient to establish a sound idea of the level of competency reached by writing pupils.

Five possible means of increasing grader reliability were compiled from synthesizing scattered data from the literature, and appear below:

1. An established set of criteria is adhered to in the grading process.
2. A ranking scale is applied in order to more accurately assess the level of competence in each of the established criteria.
3. The graders are selected on the basis of similar professional and educational background.
4. The graders meet in order to express their views concerning the criteria and the rating scale.
5. The graders practice evaluation of themes as a group, sharing their reasons for marking papers as they do.

The Jefferson County Language Arts Program, in an attempt to assess how well students were accomplishing the task of written communication, designed a program in which
graders (called judges) were selected and were to evaluate compositions according to established criteria and a rating scale.

The graders met as a unit and discussed the criteria so that a high level of agreement could be established concerning the meaning of each one. Then practice sessions were held in which each grader demonstrated exactly how he ranked a student's theme.

After the practice session, the ten judges began their work by assessing 100 student papers. They were to rank each paper with a score of one, two, three, or four, depending upon the level of proficiency exhibited.

The 100 themes were evaluated as follows: "49 papers--all 4 judges scored the same; 41 papers--1 judge varied 1 point; 10 papers--varied scoring (Townsend 1974, p. 121)."

The above study was one of many which has shown that grader reliability could be high when certain conditions, such as those presented here, were observed, and when the complexity of the evaluation was at a low or simple level.

Kehoe (1974, p. 151) stated that "... teachers who work together and train themselves in the use of a rating scale, are apt to grade essays in a more fair way or at least in a more consistent way." Compared with the single
summary mark based primarily on grader subjectivity, Kehoe's statement could be proved accurate. When a single summary mark was placed on a paper, there had been, in a sense, an unfair appraisal stated, for, rather than an appraisal based on very specific criteria known to both the student writer and the instructor, two areas alone were usually being considered, the idea expressed and the mechanics of the writing.

Those instructors who considered the expressed idea synonomous with the value of the theme might mark almost solely on that basis, and as long as the idea was commendable, how that idea was expressed, that is, in what mechanical manner, was of no consequence.

On the other hand, graders may have been concerned almost exclusively with the mechanical aspect of the written composition, and no matter how lofty or erudite the communication, a theme which had incorporated within it comma splices and misspelled words was doomed to receive a low mark. Even the handwriting on the paper could relegate the work to a low level (Davis 1974, pp. 95-96).

The Use of Criteria as Delineators of Composition Evaluation

"Teachers seem to distinguish among three degrees--high, average, and low evaluation of composition (Healy 1935, p. 53)." Composition instructors are not exceptions
to the Healy statement, and seem also to evaluate themes on three levels, high, average, and low (Arnold 1964, p. 12).

Such categorization was, for the sake of evaluation, an expedient which may have had to be employed; however, the very broad classifications could do disservice to both an instructor's evaluative attempts and to a pupil's writing practice, since narrower categories would allow both the instructor and student to see more precisely which specific steps in the sequence of the craft of writing were in most immediate need of improvement. When the specific areas were discovered, then the student could readily attack them on an individualized basis, applying whatever attention was necessary to bring about improvement (Bergman 1962, pp. 192-196).

Fairness to the pupil would be enhanced through the use of specific criteria and a set rating scale, as suggested by Kehoe (1974, pp. 146-151). The learner's work would no longer be evaluated purely on the subjective basis of one or two general aspects of the writing process, but rather on clearly established areas which could be readily discussed. Thus, while grader reliability would be increased through the use of rated criteria, so would the student's opportunity to improve his writing.
A further benefit for the learner would be the ease in which he could discover which specific part of the writing process was troubling him and to what extent. For example, if he received a consistently low mark of "2" on the criterion of organization when a "2" was on the low end of the rating scale, he could clearly see not only that he had a deficiency in that aspect of composing, but also what level his proficiency was presently on.

Knowing that he must improve in the area of "organization," the student could approach his instructor for aid, whereupon he should have received assistance in the form of tutorial guidance, a possible set of organization styles to choose from, and models to study exhibiting different types of organization. If the student consequently began to rise in the area of organization, he could see just how much improvement he had attained by noting his rating scores on that criterion.

Rather than a vast conglomeration of writing problems striking the learner simultaneously, he would be presented with only a few major important areas to improve upon at a given time, thus being allowed an opportunity to advance at his own pace, individually, and in a direction of improvement best suited to his personal needs (Combs 1966, pp. 373-376).
Before 1850 teachers of composition realized that more stimulating and understandable means for aiding pupils in the difficult task of composing clear prose were necessary (Logan 1934, pp. 486-496). One such possible stimulation might have been to clarify the process of composing by segmenting the diverse aspects of the discipline into clearly delineated categories which would facilitate both the teaching and learning process. The first half of that task had been at least partially accomplished over the years, but to date, the evaluation of the written product had been generally treated as a single unit.

Many textbooks divided the study of composition into segments from basic topic selection, through mechanics, and even to enhancement of individual writing style (Rorabacher 1956), but students' papers still received, generally, a single summary mark to denote merit, with numerous individual comments on the theme to indicate specific writing problems. The single summary mark may have been too broad to be a definitive comment to the student concerning his level of competency, and the numerous corrections may have been too confusing to offer realistic direction to the student's improvement process.

The use of criteria could, in part, avoid the stigma of the grade on a theme, placing the emphasis on instruction instead, and it had been shown as far back in time as 1913
that excessive attention to the mark given on a theme could actually impede the learning process (Barnes 1913, pp. 164-165). The criteria approach could be a motivational force because emphasis on a total grade is lessened, and replaced instead with emphasis on specific writing areas which require improvement. The instructor's attitude could change too, from concern with filling a paper with red marks of derision to one of promoting an elevation in the writing process, in which case the student may improve more rapidly and completely (Sheridan 1951, p. 322).

Criteria-centered teaching and evaluation, then, could offer a stimulus for the student to embark on a plan to better his writing skill because the threat of a grade is in part removed. Herzberg's (1952, pp. 127-131) contention that writing instruction, and writing itself, is a painful process avoided when possible could be denied at least in part by removing the fear of criticism in the form of belittling grading practices.

In the event that a grade was not awarded, and the instructor gave "... students' written work careful reading, content-centered analysis, constructive criticism, and recognition for what it communicates ...," still, in most cases the teacher evaluated a paper on confoundingly numerous grounds, then expressed the evaluation by rendering a single mark, not clearly depicting which specific
portions of the writing process were in need of improvement and to what extent (Dusel 1956, p. 327).

For the following seven reasons garnered from the literature, it seemed valid to consider the use of criteria not only in evaluating student writing, but also for elevation of instructional practice:

1. **Criteria-centered composition evaluation should** allow the student to discern the exact areas of his weakness, after which he can center his attention on those areas for improvement. "Experience tells us that not much improvement occurs unless students know what specifics they must work on to improve . . . (Townsend 1974, p. 127)."

2. **Criteria-centered composition evaluation should** allow the **instructor** to clearly describe which areas of the writing process are in most need of improvement, after which prescriptions may guide the pupil to a course of action which will lead to development of greater skill.

3. **Criteria-centered composition evaluation should not** overwhelm the student with numerous criticisms followed by a vague single summary mark; rather, the method should provide a positive approach to improving the writing skill step by step in a sequential manner, one criterion at a time until the entire process shows elevation.
4. Criteria-centered composition evaluation may then allow for a sequentially designed program to follow, and it has been shown that there is a definite "underlying developmental sequence" present in the learning of written structure (Chomsky 1972, p. 26).

5. Criteria-centered composition evaluation should allow for a much more objective approach to the consideration of student writing, and teachers themselves have admitted that a great deal of prejudice may underlie the mark a student receives on his written compositions (Guffin 1974, p. 142).

6. Criteria-centered composition evaluation may lead to an educational objectives format for studying and assessing the writing act, with all of the advantages of that system, as suggested by Mager (1962).

7. Criteria-centered composition evaluation should allow a student to progress at his own learning pace and in an individual manner, which many scholars have felt is imperative to the efficient learning of writing (Cook 1968, p. 1175). The precepts of "individualized instruction" are easily conformed to also, since the student is not held to the activities of the class, but is encouraged to operate within the specified areas which constitute his most
urgent needs. Howes (1970, p. 71) believed that such instruction based upon individual needs was more meaningful and possessed a greater potential for retention.

**Review of the Specific Criteria Selected for the Study**

Selection of specific criteria, whether for purposes of study or instruction, is, in part arbitrary. There are so many minute, yet important, aspects of the writing process that it is difficult to support the selection of one criterion over another. However, criteria for gauging writing proficiency were established early in the literature, as expressed here by Thach (1898, pp. 94-95) in the late nineteenth century:

> It is difficult to believe, at times, that many of the writers of college-entrance papers are English-Speaking boys. In the most mechanical points of execution--handwriting, spelling, punctuation--a large number are deficient to an appalling degree. They have no vocabulary; words do not appeal to them, or have for them the least significance. Any strength of thought or vividness of feeling that may occasionally struggle for utterance finds but one vent--slang. Unity or coherence of thought is seldom exhibited. Long chains of unrelated ideas are tacked together in a slackrope sentence, pages long, with a lot of 'ands,' when a lot of 'buts' would be equally exact. Paragraphing is seldom attempted, unless after the fashion of one student who systematically indented the lines in blocks of five. And, worst of all, there is rarely to be found evidence of power of thought or range of reading, such as one ought reasonably to expect in young men of seventeen or eighteen, who regard themselves as ready for college work.
Many of the criteria which have become important to an evaluation of the writing act were often mentioned peripherally in statements concerning composition in general, as may be observed above. Seely, in a 1930 (p. 235) discussion of his students' poor writing practices, named several additional aspects of composition evaluation. He registered numerous disquietudes concerning the writing ability of his pupils, complaining that they were "unable to develop their ideas; that they refuse to organize them; that they have little feeling for words; that their diction is monotonous and flaccid; that they are afraid of independent thought and reflection; . . . ." The scholar concluded that the students were, in general, not meeting the basic principles of effective composition.

If one scanned the above statement of the early twentieth century teacher of writing, he would be able to sift from it several basic criteria, many of which are actively supported in modern pedagogy as basic to effective written communication. Yet, at no time in the history of composition instruction had there been a master list of criteria which described the attributes of acceptable writing practice to the general satisfaction of scholars in the field. Lists had been developed, either by individual scholars or groups of scholars, but no collection of criteria had been accepted as a standard from which to operate.
When a large Arizona school district decided to undertake a study of its composition instruction program, it discovered that there was no established list of criteria to rely on for evaluation, and was obliged to prepare such a list.

Fortner (1974, p. 139) registered the following lament concerning the situation:

During the 1972-73 school year, all English Departments in the Phoenix Union High School System were required to make a product evaluation of all selective English courses. Writing courses could not be evaluated without a satisfactory criteria /sic/. Even so, we had no criteria by which a piece of student writing could be judged.

The Phoenix instructors met the problem of criteria selection by first having the subjects of their study compose themes, one theme to serve as a pretest device, another to serve as a posttest device. From the vast number of papers received, which represented two offerings from each student who was enrolled in English in the large district, 750 papers were selected for evaluation.

From these 750 papers, the instructors attempted to compile a list of those criteria which they considered most important to the writing act. The final list contained the following criteria (Fortner 1974, p. 141): (1) Communication of Idea - Fulfillment of Purpose, (2) Originality, (3) Organization, (4) Paragraphing, (5) Paragraph Structure, (6) Sentence Structure, (7) Word Choice, (8) Punctuation/Capitalization, (9) Spelling, (10) Legibility.
The ten criteria selected by the Phoenix Union district do not represent in any way, a standardized list. When a New York English faculty decided to perform a like study in which it, too, had to compile a list of criteria, it arrived at the following (Kehoe 1974, p. 147): (1) Purpose, (2) Organization, (3) Content, (4) Diction, (5) Sentences, (6) Grammar, (7) Mechanics.

After the selection of criteria, the list was often used to different purpose, again indicating a lack of professional agreement extant in the discipline. Both the Phoenix and New York schools rated their criteria on a scale; however, there was one major difference in the method of rating. The New York system agreed that all criteria were of equal importance; but the Phoenix district used weighted measures. The New York system awarded points from one to five on all criteria from purpose to mechanics.

The Phoenix system fixed its points according to criterion merit, crediting communication of idea with an upper score of nine points, highest attainable in the zero to nine scale, while next in weight was originality and organization, both earning up to seven points. Paragraphing and paragraph structure were assigned values from zero to six. Sentence structure ranged from zero to five points. Word choice was worth as much as five points, and punctuation and spelling from zero to three points. Legibility could attain one of three points, from zero to two.
Thus, while the selection of criteria itself has remained in the stage of arbitrary choice based upon the practices of the various evaluators, there has been further disparagement in the value of each selected criterion, making the process still more arbitrary in nature. Because of this situation, in part, modern scholars still may believe that the teacher should be able to simply recognize good writing in a general, almost artistic manner, from his intuition, experience, and knowledge. He must be "... able to tell good writing when he sees it and to help the student to do so, or how will the student come to evaluate his own writing?," asked Blake (1970, pp. 7-8).

The problem has been, quite clearly, that some of the criteria are subjective in nature, while others are much more objective. For example,

On the level of highest objectivity is the criterion of correctness: grammar, spelling, and mechanics. Next comes matters of organization of ideas along with paragraph structure and development. On the least objective level, what the statistical evaluators would call the subjective level, are the criteria of originality, vitality, and significance of ideas, as well as of the mature and reasoned development of ideas (Koclanes 1961, p. 254).

In spite of the obvious truth in the Koclanes statement, criteria may still be manipulated with a great degree of accuracy when a scale is used, as shown by studies. When certain factors are observed, such as grader compatibility and practice, there is a possibility of high correlation.
Supported by this observation, the researcher may select criteria by investigating the literature and choosing those which have appeared most often as selected by the scholars.

As for the number of criteria used in an evaluation or instructional venture, the literature has shown further variance. Some studies have employed as many as ten, as did the Phoenix Union model, while others have listed as few as five, or less, as in the present study. The number of criteria used for instruction or evaluation revolves around the pedagogical and heuristic beliefs of the experimenters.

Too many criteria, the author believes, may tend to cloud the issue when evaluating student efforts, confusing both the evaluator and the pupil. Scholars in the early part of this century recognized that instructors often attempted to entertain too many areas at one time, thus confounding the basic precepts of a program. The students became mired in a confusing plethora of objectives, losing sight of the primary purpose of the writing program, to express thoughts effectively and clearly in written form (Leonard 1916, pp. 598-604).

While, for the purpose of the present study, the criteria were considered as independent of one another, the writer recognized that this was, at least in part, an
artificiality adhered to for reasons of facility. All the criteria coexisted as parts of the whole, each having a specific and important value in the end written product. Mechanics, for example, enjoyed no greater value than communication of idea, for it was, in part, responsible for part of the written communication in itself (Vivian and Jackson 1967, p. 356). Below, each criterion has been described separately as it was treated in the literature.

Organization

Christensen (1965, pp. 144-156) explained that writing is a process of organized addition, and that within the act of composition is the concept of order, sequence, and development. There is a main idea, and that main idea is modified in an organized fashion through addition.

Irmscher (1969, p. 29) considered writing a formal process much more complex and less spontaneous than speaking, and therefore, in need of deliberate and careful organization. A major problem evident in student themes over the years has been a lack of careful organization, and this element was perhaps included more than any other in compilations of writing criteria (Saalbach 1958, p. 505).

Students need, "... first of all, instruction in the ordering of ideas they have, said Fichtenau (1965, p. 721)." Researchers have continued to find that a major
problem attending beginning writing proficiency has been an inability for the student to isolate a given idea, or theme, and then develop that main idea carefully and consciously (Rankin and Fleming 1974, p. 47). While the problem was associated with students of the primary grades in the Rankin-Fleming study, there has been good evidence that older students remain inept in the same area, having difficulty in defining a main idea and organizing a writing plan which would communicate that idea (Crosby and Estey 1968, p. 55). Many texts designed to supplement the instruction of college composition have expressed a need for students to clearly organize main thoughts through preparation of a formal outline in order to insure written sequential presentation (Arntson 1961).

Piaget (1961, pp. 275-281) suggested that organization was an important part of the thought process itself, and that understanding was served best through one communicator carefully organizing the information to be passed on to a receptor. While the psychologist was concerned primarily with speech communication, his observation may be applied as well to written communication.

Grommon (1959, p. 17) stated that "A good student writer will demonstrate a high level of proficiency in organization, combining clear sentences in well-shaped paragraphs and arranging these in order clear to the reader
as well as to himself." The criterion of organization has been expressed as a part of the conveyed meaning of the written product, with changes in order of presentation altering the communication.

Types of organization have been many, and the specific plan decided upon by the writer may well depend upon the communication intended and the audience (Rorabacher 1956, p. 7). Crosby and Estey (1968, p. 56) offered the idea that "Organization provides a path by which the writer can lead his reader through a maze of information and arrive at the conclusion desired."

Thus, an evaluator must bear in mind that no one type of organization may be considered standard. "The arrangement or organization of the material of a communication must be governed by its purpose. That is why no one formula for organization will work on all occasions (Mills 1953, p. 20)." The type of organization, then, was not considered the important factor. The important factor was whether or not the organizational foundation of the essay served to further the success of the paper. As Francis (1965, p. 118) succinctly pointed out, "Order is a prerequisite of meaning."

Mechanics

Over the past decade the teaching of mechanics has been relegated to a low position in the total composition
instruction process. Probably, this was due to a mistaken coupling of mechanics instruction with grammar instruction. Research had shown that the teaching of grammar in an isolated form had little to do with writing improvement. In fact, in some cases grammar instruction was credited with negative value, as Braddock et al. (1963, pp. 37-38) stated,

In view of the widespread agreement of research studies based upon many types of students and teachers, the conclusion can be stated in strong terms: the teaching of formal grammar has a negligible or, because it usually displaces some instruction and practice in actual composition, even a harmful effect on the improvement of writing.

This viewpoint, however accurate, may have been responsible for the popular concept that it was only the idea in writing that was of meritorious consideration, and as long as that idea could be clearly understood, even if the audience had to ponder in order to grasp it, the major task of writing had been accomplished. This concept of mechanics in writing was, in the opinion of the investigator, a faulty one. Mechanics should be considered important in teaching the writing process, for it has remained a criterion which, along with others, aids in the precision of the communication.

If the basic unit of written understanding was to be the sentence, as many scholars have propounded, then how that sentence was composed, with what marks of punctuation, must
have been a decided and distinct part of the communication (Ohmann 1966, p. 261). Graves (1972, p. 701) considered the mechanical aspect of a written communication to be highly significant, and interpreted Christensen's ideas on form to be a "conscious control" which could change the meaning of a piece of writing.

Mechanics, unlike grammar, could not exist in isolation from the writing process, but was to be interwoven in the written communication in a consciously controlled fashion to convey specific meaning.

Mills (1953, p. 24) discussed the value of mechanics and structure as a device of communication by supporting the belief that the form was not as important as the idea, but that the form could control the conveyance of the idea. He said,

Students must be made to see that mechanics is theoretically the least important aspect of the process of communication—that purpose and material and organization and expression are the really functional elements in a piece of writing and that grammatical purity without these other elements can achieve nothing. But they should be made to see also how egregious failures in grammatical appropriateness can undermine and destroy the effectiveness of all these other elements—how one misspelled word can destroy the good impression of an otherwise satisfactory letter of application, how a confusion in sentence structure or agreement can almost totally obscure the meaning of a statement, how an amusing grammatical blunder can dissipate the moving effect of an otherwise persuasive argument.
As Vivian and Jackson (1967, p. 356) forcefully stated: "A careful writer may sometimes take liberties with conventional punctuation, but he does so only when he knows that such liberties will not mislead or confuse his readers. He keeps in mind that his reason for punctuating what he writes is to make his meaning clear."

Vocabulary

The writer's vocabulary has been considered the sum total of words he has used to express himself. As this vocabulary increased, so did the opportunity to have the proper word available for effective use at a given time. An ability to present the most effective word for the particular circumstances had to do with diction, as defined by Vivian and Jackson (1967, p. 304): "Diction is the selection of words to express ideas. Good diction is the choice of words effective and appropriate in meaning, as well as suited to the subject, audience, and occasion."

While there was never a perfect correlation of 1.00 in communication, a speaker or writer never conveying 100 percent his exact meaning, a higher correlation could be attained through the use of the correct words; hence, the word was considered by many the basic tool of expression, both cognitive and emotive (Allen 1942, p. 179).

Often words were categorized according to the emotion or idea they represented, such as "concrete,"
"abstract," "action," "descriptive," or other (Perkins 1964). Without a command of vocabulary, written expression usually failed. As Frazier (1964, p. 436) said, "Immature writing is typically general or imprecise in its language. Deliberate efforts to affect the use of words in writing may be worthy of more attention."

Because of its importance to the writing process, both in teaching and evaluating, vocabulary deserved a place in any list of criteria. In the minds of some scholars, words were more than expressors of the mind; they in fact might, as Aldous Huxley suggested, come first in the thought process, leading into a particular means of assimilating an experience. He said, "Words form the thread on which we string our experiences (Watkins and Knight 1966, p. 180)." Huxley perhaps overstated the value of vocabulary; however, though man does not, it has been believed, think totally in words, but also in emotions, instincts, subconscious and unconscious patterns, still the word has been the primary unit of understanding and expression, and "A student can be taught at least to write and speak so that words do not get in the way of his thoughts (Macrorie 1951, p. 385)."

Communication of Idea

Vivian and Jackson (1967, p. 1) have stated that "The basic reason why anyone studies composition is so that
he may learn to communicate with maximum effectiveness."

Further, communication has been considered the end product of a piece of writing, the main element, and the result of applying all of the criteria of writing with effectiveness. At the same time, a problem exists, for abundant communication of an idea may have taken place in the midst of highly inadequate composition practices.

The idea may have been conveyed in spite of ineffective organization, vocabulary, mechanical structure, or logical development. The reader may have been able to extract from the writing, with cautious deciphering, the major points in spite of, not because of, the writer's presentation. In such a case, the paper, although communicative, was a failure and an inadequate sample of writing. The idea may have been expressed, but the manner of expression may indeed have caused the reader to consider an opposite viewpoint from that intended by the author. The reader may have decided that the idea was as inadequate as its expression. One may assume that under certain circumstances the idea can be no better than its manner of presentation.

When Larson (1974, p. 81) stated that "The process of writing should lead to communication . . . ," he was concurring with Skinner (1969, pp. 257-260, 266) that breeches in compositional construction could result in a loss of
written expression, since structure may, of itself, communicate. This, in effect, supported the idea that all the criteria were interrelated and dependent upon each other, and it also supported the autonomy of each criterion for no one was placed above another in value, not even communicative value. Communication of idea, then, has remained separate as one category, in the present study, and given no additional weight when used as a measuring device for composition evaluation.

Logical Developmental Sequence

Although combined for reasons of the study, logical developmental sequence may be divided into (1) logic and (2) sequence for discussion.

Logic. While most scholars would agree that few disciplines are, of themselves, mind expanding or conducive to elevation of the thought processes, it has been believed by many that writing could enhance cognitive ability because of its aid in developing logical patterns of thinking. Vivian and Jackson (1967, p. 11) have stated that:

Good writing depends upon clear thinking; we cannot write well unless we have something sensible to say. Conversely, clear thinking depends upon the careful use of language; we cannot think clearly unless we formulate our ideas precisely in words. Essential to logical reasoning is the ability to say accurately what we mean.

Psychologist Jerome Bruner (1966) had referred to composition as the calculus of thought, insisting that a
writer was only as lucid and forceful as his ability to clearly and forcefully think. Zoellner (1969, p. 289), to add to the list of those who correlated the writing process with the thought process, stated that "... faulty thought always comes first as cause, and faulty writing second as effect." Hook (1965, p. 226) stated that "Composing is thinking." Quoting another authority, Blake (1970, p. 196) believed that "Expository writing is the result of disciplined thinking. It is a complex activity involving the skills of creating syntactic constructions, a knowledge of words and their meanings, the ability to recreate observations in writing, and the ability to follow through thought processes logically." Zoellner (1969, p. 269) again insisted that: "The pervasively defining instrumental metaphor in our teaching of composition and rhetoric can be easily and succinctly stated: it asserts that the written word is thought on paper." Finally, Hodges and Whitten (1967, p. 254) combined both logic and sequence by stating that effective prose must possess logical order.

**Sequence.** The process of sequential step-by-step development of the logically arrived at components of a written product was not only important because "All good writing, absolutely all, is both coherent and developmental (Scholes and Klaus 1972, p. 5)," but also because of the added retention value and communicative forcefulness of sequential design (Means 1974, p. 160).
While the benefit of sequential design was to bring both forcefulness and retention to the writing itself, there was also heuristic value in "... establishing a systematic, step-by-step procedure for teaching the process of writing (Larson 1974, p. 80)." For, as Mills (1953, p. 19) suggested, "The most fruitful developments in pedagogy in our field are all based on concepts of process." Further, "Teachers should have a rationale for teaching composition and be able to explain it to students ... (Rankin and Fleming 1974, p. 48)." The rationale teachers of composition proffer should be designed in such a way as to make clear a logical, sequential design which students may readily learn from.

Summary

This chapter has been concerned with a literature review of the component parts of the study, being, primarily, the effect of theme topic selection on written compositions, rating scales in the evaluation of composition skills, the employment of graders and their reliability in evaluating written composition, the use of criteria as delineators of composition evaluation, and a review of the specific criteria selected for the study.

Topic selection, it has been shown, can affect many aspects of the writing act. Therefore, it has been important to consider how topics were selected and which topics
comprised the subjects for any themes which would consequently be evaluated.

Rating scales were developed early in the twentieth century, modified, and employed in various ways in composition evaluation, which, some scholars believed, would upgrade both the level of grader agreement and fairness to the student's writing.

Scholars had recognized that much of the evaluation rendered on student writing had been, at best, a subjective commentary which would change capriciously as the grader's whims changed. This condition may have caused the discipline to consider less subjective means of criticizing composition skills, such as the use of criteria.

The employment of criteria for evaluation of composition skills was considered a possible solution for a prevalent teacher tendency to grade in three major areas, high, average, and low. More precise evaluation, it was shown in the literature, would be possible through an application of specific criteria. Seven attributes of criteria-centered evaluation were generated from the literature.

A review of the five criteria which were selected for the study based in part on their prevalence in the literature composed the last section of Chapter II.
CHAPTER III
PROCEDURES

This chapter is concerned with the various procedures which gave direction to the course of the research.

The Research Design

The one-group, pretest-posttest design, while widely considered a quasi-experimental procedure, was found to be of practical application in the present study, which was planned to initiate cursory investigative processes in a composition course, in this case The University of Arizona Model Freshman Composition Program. The design was one of the only feasible models due to institutional factors and conditions extant at the time of the study. Further, certain of the design’s limitations were overcome through specific conditions peculiar to the present study.

The investigation did not pose a true hypotheses testing condition; however, because of the design, an inherently positive factor existed, which was experimental isolation. The study was of short duration, slightly less than one academic semester, or approximately four months, thus the effect of pupil maturation was lessened. Furthermore, it would be possible to determine Grader-awarded point
growth on each of the five criteria of the study, from organization to logical developmental sequence, with the present design.

There was however, a negative factor in that it would not be clearly discernible as to whether X or history, maturation, statistical regression, or other extraneous variables were responsible for the variance between $0_1$ and $0_2$ ratings; however it must be remembered that no design, including the most stringent experimentally controlled paradigm, has produced results which could be guaranteed totally free of extraneous variables, due in part to the fact that the sum total of man's underlying reasons and causes for behavioral change have yet to be isolated.

History and maturation will have taken place during the course of any study, and must be considered as factors in resulting changes in subjects. The University of Arizona Model Freshman Composition Program had considered this aspect as internal to its program, intending not only to enhance the students' ability to perform at a higher level of competence in the writing act, but also intending to affect promotion in thought patterns and ideas through planned classroom discussions and activities.

**The Population**

The population consisted of the twenty-two sections of freshmen students enrolled in The University of Arizona
Model Freshman Composition Program during the second semester session of 1973-74. The student population represented a homogeneous group assigned to each given section purely on the basis of physical reason, such as class schedule, and never for reasons of academic selectivity.

The sample consisted of three classes extracted at random from the twenty-two sections comprising the population. Instructors had the prerogative of refusing inclusion in the study, which two exercised. Hence, the investigator was obliged to select again from the population in order to arrive at the three sections which would constitute the subjects for the study. There were seventy-four pupils total for the three sections.

After administration of the $O_2$ theme, it was found that fourteen of the students did not write both an $O_1$ and an $O_2$ paper. These fourteen students were then eliminated from the sample, leaving sixty subjects who had completed both themes. Entirely at random, ten sets of papers were extracted, leaving the final sample group at fifty subjects.

**Assurance of Student Anonymity**

The Human Subjects Committee of The University of Arizona granted permission for the present study on the grounds that students and their performances remain entirely anonymous. This was rigidly observed through the assignment of four-digit numbers which were selected from a
telephone book and attached to each student's theme. The number was the only means of identification for the experimenter, and a master list of names and numbers was held only by the course instructors and later destroyed.

**Collection of Student Data**

Eight items comprised the student information list for the study. These were: (1) sex, (2) grade point average in high school English, (3) A.C.T. score in English, (4) age, (5) bilingualism--non-bilingualism, (6) attendance in The University of Arizona Model Freshman Composition Program, (7) economic rank of the student's family, and (8) high school graduation rank. The student data were collected by three means, a questionnaire, a direct search of the records on file in the Registrar's Office of The University of Arizona, and the class instructors' record books. Items number four, five, and seven were compiled from the questionnaire, a copy of which may be found in Appendix A. Items number two, three, and eight were collected directly from the files of the Office of the Registrar, and items number one and six were retrieved from the record books of the instructors.

Student anonymity was preserved by using the master list to compile matriculation numbers only, after which all data was transferred to coded information sheets which bore only student identification numbers.
Criteria Selection

Criteria were selected on three bases, the review of the literature and a detection of modality therein, a review of the textbook used in The University of Arizona Model Freshman Composition Program, *Effective Writing*, 4th edition, New York: Holt, Rinehart and Winston, Inc., 1971, by Robert Hamilton Moore, and a review of the educational objectives as prepared by Dr. Charles Davis for The University of Arizona Model Freshman Composition Program.

There was a high level of agreement among the three sources above concerning the inclusion of the five criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence, which were selected for evaluating the variance between pretest and posttest themes.

The Test Instrument

The test instrument, or evaluation sheet, as it was designated for the study, consisted of the five criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence, rated on a five-point Likert-type scale with a low of one and a high of five. A sample of this instrument may be found in Appendix B.

Each Grader was presented also with a list of definitions which explained the investigator's concept of each criterion so that a level of Grader agreement concerning
the meaning of each could be agreed upon. A sample of this list of criteria definitions may be found in Appendix C.

Selection of Graders

Five persons were selected to act as Graders, or evaluators, of the \( \theta_1, \theta_2 \) themes. They were chosen on the basis of professional experience in writing for publication and experience as instructors of writing. All five were teaching college composition courses during the study and were scattered widely geographically. The five were, for the purpose of the study, labeled Grader A, Grader B, Grader C, Grader D, and Grader E.

Establishment of Reader Reliability Coefficient Using the Pearson Product Moment Correlational Technique

Ten pre-study sample themes were distributed to the five Graders in order to establish their level of agreement on employing the test instrument. Eight of the themes were selected at random from a previous class of freshman students who were enrolled in The University of Arizona Model Freshman Composition Program in 1972. One of the two remaining themes was selected from a small publication on the basis of concise clarity and high level of communication. The final theme was selected from a group of high school essays on the basis of its lack of acceptable form and absence of a clearly defined and communicated idea.
The ten themes were typed faithfully from the original forms and distributed to the Graders with the rating sheet instrument attached to each. The Graders rated and returned the ten essays to the experimenter for analysis. Table 1 below expresses the results of the Pearson product moment correlation for each variable pair, and one summary figure.

Table 1. Pearson Reader Reliability Correlation Coefficient for Pre-Study Sample Themes.

<table>
<thead>
<tr>
<th>Variable Pair</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>.56</td>
</tr>
<tr>
<td>AC</td>
<td>.73</td>
</tr>
<tr>
<td>AD</td>
<td>.65</td>
</tr>
<tr>
<td>AE</td>
<td>.41</td>
</tr>
<tr>
<td>BC</td>
<td>.61</td>
</tr>
<tr>
<td>BD</td>
<td>.71</td>
</tr>
<tr>
<td>BE</td>
<td>.60</td>
</tr>
<tr>
<td>CD</td>
<td>.71</td>
</tr>
<tr>
<td>CE</td>
<td>.42</td>
</tr>
<tr>
<td>DE</td>
<td>.55</td>
</tr>
</tbody>
</table>

Reader reliability coefficient: .61
While a reader reliability coefficient of .61 may be considered less than desirable, it must be observed that this correlation represented the Graders' attempts to match five criteria on a five-point scale, unlike many previous studies in which simple ranking of essays resulted in reader reliability coefficients above .80.

By inspection of the rating sheets it was determined that each Grader successfully detected the single paper included for its high level of performance, and the single paper included for its low level of performance, although in no case did any two Graders rate these papers in exactly the same manner on all five criteria.

**Topic Selection**

Topics were selected on the basis of modality from questionnaires presented to four different groups: two classes of freshman composition students, neither of which would be included in the study, one class of seniors at the college level, one class of graduate students, and one adult evening class. The ten topics which received the most frequent listing were the following: the energy crisis, government corruption, the drug problem, impeachment, American social structure, abortion legality, ecology problems, present value of a college education, religion in America today, and the generation gap. These topics were most popular when
the students were asked to fill three blanks with subjects they would consider important ones for writing assignments. A sample of the above questionnaire may be found in Appendix D. Half of the ten topics selected from the aforementioned questionnaires appeared on the $O_1$ theme assignment sheet, and half appeared on the $O_2$ theme assignment sheet.

**Process for Administering Pretest and Posttest Themes**

The $O_1$ and $O_2$, pretest-posttest themes were administered in the following manner: the students were asked to write for fifty minutes on one of five topics appearing on a written assignment sheet which was distributed by the investigator. A sample of the $O_1$ theme assignment sheet may be found in Appendix E, and a sample of the $O_2$ theme assignment sheet may be found in Appendix F.

**Process for Scoring the Pretest, Posttest Essays**

All papers of the $O_1$, $O_2$ groups which were reviewed and rated by the Graders were typed faithfully from the in-class handwritten originals in order to insure against bias due to penmanship (Davis 1974, p. 95). After a faithful typing of the original $O_1$, $O_2$ themes, the pairs were thoroughly mixed, with no possible means of identifying $O_1$ or $O_2$ papers, and delivered to the five Graders, rating sheets
attached. The Graders rendered the appropriate ratings and returned the themes to the experimenter for analysis.

**Analysis of the Data**

The data were analyzed by the following measures:

1. A simple compilation of raw data to determine percentage increases in Grader-awarded points on each of the five study criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence;
2. A correlated t-test of the difference between two means in order to establish the statistical level of significance between the pretest and the posttest;
3. A correlated t-test of the differences between two means in order to establish the statistical level of significance between the pretest and posttest concerning each of the eight student variables: sex, grade point average in high school English, A.C.T. score in English, age, bilingualism--non-bilingualism, attendance in The University of Arizona Model Freshman Composition Program, economic rank, and high school graduation rank;
4. A step-wise multiple linear regression technique for the purpose of accounting for the variance on the posttest in relation to each of the eight student variables (above) and the pretest; and
5. A Pearson product moment correlation measure to establish reader reliability on the pretest and on the posttest.
Summary

The research design of the study was the one-group pretest-posttest paradigm, employing student-composed themes for the $0_1, 0_2$ analyses. The population encompassed the twenty-two sections of Freshman Composition at The University of Arizona during the second semester of the 1973-74 academic year. Three sections were selected at random from the population to become the sample for the study.

Student anonymity was insured by the experimenter through the employment of identification numbers and the omission of names on all study materials. The student variables, which included: sex, grade point average in high school English, A.C.T. in English, age, bilingualism--non-bilingualism, attendance in The University of Arizona Model Freshman Composition Program, economic rank, and high school graduation rank were collected by means of a questionnaire, records on file in The University of Arizona Registrar's Office, and the classroom instructors' record books.

The study criteria were five: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence. These were selected on the basis of a search of the literature, and a study of the textbook and syllabus which served to guide the learning functions of The University of Arizona Model Freshman Composition Program.
Student themes were examined by five Graders who had been selected on the basis of experience in professional writing and composition instruction. The Graders judged $0_1$, $0_2$ student themes with a five-point Likert type rating instrument based upon the five study criteria cited above. Graders' initial reader reliability coefficient was established at .61 by means of a Pearson product moment correlation test of ten pre-study sample themes.

The topics for the $0_1$, $0_2$ student themes were selected by mode from a student questionnaire which asked various pupils to list subjects which they considered appropriate for writing essays. The ten most often mentioned topics were selected, half of which were attached to the $0_1$ theme assignment, half to the $0_2$ theme assignment.

The subjects of the study wrote two fifty-minute essays, one early in the semester, one late in the semester. After both $0_1$, and $0_2$ themes were collected by the experimenter they were typed faithfully from the handwritten originals, mixed thoroughly so that $0_1$ themes could not be distinguished from $0_2$ themes, and distributed to the five Graders for their ratings based on the five criteria, five-point scale.

The data were analyzed by five means: (1) raw score compilation to determine percentage gains in Grader-awarded
points on each of the five study criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence; (2) a correlated t-test of the difference between two means in order to discover the statistical level of significance between the pretest and the posttest; (3) a correlated t-test of the differences between two means for the purpose of establishing the statistical level of significance between the pretest and posttest based upon the eight student variables: sex, grade point average in high school English, A.C.T. score in English, bilingualism--non-bilingualism, attendance in The University of Arizona Model Freshman Composition Program, economic rank, and high school graduation rank; (4) a step-wise multiple linear regression technique performed in order to account for the variance on the posttest in relation to each of the eight student variables (above) and the pretest; and, (5) a Pearson product moment correlation measure to establish reader reliability on the pretest and on the posttest.
CHAPTER IV

FINDINGS

The basic findings of the present study were divided into the following five areas: (1) percentage of increase from pretest to posttest based upon each of the five study criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence, in terms of accumulated Grader-awarded raw points, (2) results of a correlated t-test of two means for the purpose of establishing statistical significance between the pretest and posttest, (3) results of a correlated t-test of the differences between two means in order to establish the statistical level of significance among the eight student variables: sex, grade point average in high school English, A.C.T. score in English, bilingualism--non-bilingualism, attendance in The University of Arizona Model Freshman Composition Program, economic rank, and high school graduation rank, based on the pretest-posttest changes, (4) results of a step-wise linear regression analysis of the variance exhibited on the posttest dependent variable in association with the eight student data, from sex to high school graduation rank, as expressed above, and the pretest functioning as independent
variables, (5) results of the Pearson product moment correlation executed to establish the reader reliability coefficient for the five Graders based on their pretest and posttest ratings.

### Percentage of Increase on Each Individual Study Criterion in Terms of Accumulated Points

There was an increase in number of points awarded by the five Graders based on the five-point Likert-type scale on each of the five study criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence. Calculation based on 50n revealed the percentage of this increase on each criterion, as illustrated in Table 2. The null hypothesis, \( H_0 \), was rejected because of a clearly definitive and discernible advance in Grader-awarded points based upon the five study criteria as described above. The results of the raw score data compilation decidedly permitted the rejection of \( H_0 \), which had been stated as follows: there will be no difference between the pretest and posttest themes in terms of the criteria.

The compilation of raw scale points in each criterion revealed a thirty-five percent increase between \( O_1 \) and \( O_2 \) Grader-awarded scale points on the organization criterion, a sixty-one percent increase between \( O_1 \) and \( O_2 \) Grader-awarded scale points on the mechanics criterion, a five percent increase between \( O_1 \) and \( O_2 \) Grader-awarded
Table 2. Grader-Awarded Raw Point Percentage Gains Per Individual Study Criterion.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Raw Point 01 Total</th>
<th>Raw Point 02 Total</th>
<th>Student Criterion Average for 01</th>
<th>Student Criterion Average for 02</th>
<th>Percent of Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organization</td>
<td>739*</td>
<td>994*</td>
<td>14.8</td>
<td>19.9</td>
<td>35%</td>
</tr>
<tr>
<td>Mechanics</td>
<td>530*</td>
<td>853*</td>
<td>10.6</td>
<td>17.1</td>
<td>61%</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>919*</td>
<td>965*</td>
<td>18.4</td>
<td>19.3</td>
<td>5%</td>
</tr>
<tr>
<td>Communication of Idea</td>
<td>611*</td>
<td>863*</td>
<td>12.2</td>
<td>17.3</td>
<td>41%</td>
</tr>
<tr>
<td>Logical Developmental Sequence</td>
<td>752*</td>
<td>992*</td>
<td>15.0</td>
<td>19.8</td>
<td>32%</td>
</tr>
</tbody>
</table>

*Maximum total possible points on each criterion 1250

n=50
scale points on the vocabulary criterion, a forty-one percent increase between 0₁ and 0₂ Grader-awarded scale points on the communication of idea criterion, and a thirty-two percent increase between 0₁ and 0₂ Grader-awarded scale points on the logical developmental sequence criterion.

Results of the Correlated t-test of Two Means for Establishing Statistical Level of Significance for Pretest and Posttest

A correlated t-test of two means rendered a two-tailed probability level beyond the .001 range for the pretest-posttest based on 50n and forty-nine degrees of freedom, insuring a powerful test for linear regression. A correlation of .69 was established on the basis of the pretest-posttest relationship. The results of the present correlated t-test may be examined in illustrated form in Table 3.

Results of the Correlated t-test of Differences Between Two Means Based on the Eight Student Variables

The statistical level of probability for each of the eight student variables: sex, grade point average in high school English, A.C.T. in English, age, bilingualism--non-bilingualism, attendance in The University of Arizona Model Freshman Composition Program, economic rank, and high school graduation rank, in accord with their increases on the posttest, was established through a correlated t-test of significant differences for two means.
Table 3. Correlated t-test of Two Means for the Pretest-Posttest.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>69.18</td>
<td>24.78</td>
<td>13.19</td>
<td>1.87</td>
<td>13.29</td>
<td>.69</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Posttest</td>
<td>93.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n=50
Each student variable was divided into two levels in order to present a compatible format for the t-test design. These levels were: sex, male on the first level, female on the second level; grade point average in high school English, A and B on the first level, C, D, and F on the second level; A.C.T. in English, twenty points and above on the first level, nineteen or fewer points on the second level; age, twenty years or more on the first level, nineteen or fewer points on the second level; bilingualism--non-bilingualism, English-speaking only on the first level, English plus any other language on the second level; attendance in The University of Arizona Model Freshman Composition Program, missed five days or less on the first level, missed more than five days on the second level; economic rank, $8,000 or above on the first level, below $8,000 on the second level; and high school graduation rank, fiftieth percentile and above on the first level, and below the fiftieth percentile on the second level. Table 4 illustrates the division of the eight student variables as they were arranged for processing in the correlated t-test.

A summary of the results of the t-test analysis is presented in Table 5. The levels of probability established by each of the eight student variables were as follows: sex, male--beyond .001, female--.029; grade point average in high school English, A, B--beyond .001, C, D, F--.001;
Table 4. Delineation of Student Data Variables for the Correlated t-test.

<table>
<thead>
<tr>
<th>Description</th>
<th>Discrimination</th>
<th>Number of Students In Group</th>
<th>Percent of n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>male</td>
<td>33</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>G.P.A. in h.s.</td>
<td>A,B</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>C,D,F</td>
<td>26</td>
<td>52</td>
</tr>
<tr>
<td>A.C.T. in English</td>
<td>20+</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>19-</td>
<td>39</td>
<td>78</td>
</tr>
<tr>
<td>Age</td>
<td>20+</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>19-</td>
<td>45</td>
<td>90</td>
</tr>
<tr>
<td>Bilingualism--non-bilingualism</td>
<td>English only</td>
<td>40</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>English + other</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Attendance</td>
<td>missed 5 or less</td>
<td>33</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>missed more than 5</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>Economic rank</td>
<td>$8,000 +</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>-$8,000</td>
<td>28</td>
<td>56</td>
</tr>
<tr>
<td>H.S. graduation rank</td>
<td>50th percentile</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>and above</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>below 50th</td>
<td>37</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>percentile</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n=50
Table 5. Correlated t-test of Two Means on the Eight Student Variables.

<table>
<thead>
<tr>
<th>Test</th>
<th>Variable</th>
<th>No. of Cases</th>
<th>Mean Diff.</th>
<th>Mean</th>
<th>S.D.</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>Corr.</th>
<th>2-tail Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>Sex, Male</td>
<td>33</td>
<td>26.49</td>
<td>67.24</td>
<td>19.58</td>
<td>3.41</td>
<td>11.23</td>
<td>.74</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td></td>
<td>93.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>Sex, Female</td>
<td>17</td>
<td>21.47</td>
<td>72.94</td>
<td>14.29</td>
<td>3.47</td>
<td>7.28</td>
<td>.53</td>
<td>.029</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td></td>
<td>94.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>G.P.A., A, B</td>
<td>24</td>
<td>21.54</td>
<td>71.88</td>
<td>18.22</td>
<td>3.72</td>
<td>8.94</td>
<td>.78</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td></td>
<td>93.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>G.P.A., C,D,F</td>
<td>26</td>
<td>27.77</td>
<td>66.69</td>
<td>17.82</td>
<td>3.49</td>
<td>10.19</td>
<td>.63</td>
<td>.001</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td></td>
<td>94.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>A.C.T. 20+</td>
<td>11</td>
<td>15.91</td>
<td>82.00</td>
<td>17.69</td>
<td>5.33</td>
<td>4.98</td>
<td>.82</td>
<td>.002</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td></td>
<td>97.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5, Continued. Correlated t-test of Two Means on the Eight Student Variables.

<table>
<thead>
<tr>
<th>Test</th>
<th>Variable</th>
<th>No. of Cases</th>
<th>Mean</th>
<th>Mean Diff.</th>
<th>S.D.</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>Corr.</th>
<th>2-tail Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>A.C.T. -20</td>
<td>39</td>
<td>65.56</td>
<td>27.28</td>
<td>16.59</td>
<td>2.66</td>
<td>13.24</td>
<td>.63</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>92.85</td>
<td>10.20</td>
<td>1.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>Age, 20+</td>
<td>5</td>
<td>66.40</td>
<td>20.20</td>
<td>9.81</td>
<td>4.39</td>
<td>4.25</td>
<td>.59</td>
<td>.290</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>86.60</td>
<td>12.93</td>
<td>5.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>Age, 19-</td>
<td>45</td>
<td>69.49</td>
<td>25.29</td>
<td>18.76</td>
<td>2.80</td>
<td>12.62</td>
<td>.72</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>94.78</td>
<td>10.08</td>
<td>1.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>English only</td>
<td>40</td>
<td>69.13</td>
<td>23.85</td>
<td>17.69</td>
<td>2.80</td>
<td>12.26</td>
<td>.73</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>92.98</td>
<td>11.01</td>
<td>1.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>English + other</td>
<td>10</td>
<td>69.40</td>
<td>28.50</td>
<td>20.26</td>
<td>6.41</td>
<td>5.46</td>
<td>.64</td>
<td>.047</td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>97.90</td>
<td>7.59</td>
<td>2.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5, Continued. Correlated t-test of Two Means on the Eight Student Variables.

<table>
<thead>
<tr>
<th>Test</th>
<th>Variable</th>
<th>No. of Cases</th>
<th>Mean</th>
<th>Mean Diff.</th>
<th>S.D.</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>Corr.</th>
<th>2-tail Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>Attendance missed 5 or less</td>
<td>33</td>
<td>71.03</td>
<td>17.68</td>
<td>3.08</td>
<td>10.80</td>
<td>.78</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>92.97</td>
<td>10.21</td>
<td>1.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>Attendance more than 5 missed</td>
<td>17</td>
<td>67.52</td>
<td>19.19</td>
<td>4.19</td>
<td>8.85</td>
<td>.65</td>
<td>.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>28.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>96.05</td>
<td>10.49</td>
<td>2.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>Income $8000 +</td>
<td>22</td>
<td>70.73</td>
<td>24.68</td>
<td>19.09</td>
<td>7.89</td>
<td>.69</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>95.41</td>
<td>8.20</td>
<td>1.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>Income below $8000</td>
<td>28</td>
<td>67.96</td>
<td>24.86</td>
<td>17.38</td>
<td>10.81</td>
<td>.71</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td></td>
<td>92.82</td>
<td>12.09</td>
<td>2.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5, Continued. Correlated t-test of Two Means on the Eight Student Variables.

<table>
<thead>
<tr>
<th>Test</th>
<th>Variable</th>
<th>No. of Cases</th>
<th>Mean</th>
<th>Mean Diff.</th>
<th>S.D.</th>
<th>Standard Error</th>
<th>t-Value</th>
<th>Corr.</th>
<th>2-tail Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest</td>
<td>Grad. rank</td>
<td>13</td>
<td>75.46</td>
<td>18.15</td>
<td>23.23</td>
<td>6.44</td>
<td>5.16</td>
<td>.87</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>50%+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td>93.62</td>
<td></td>
<td>14.78</td>
<td>2.56</td>
<td>12.97</td>
<td>.58</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pretest</td>
<td>Grad. rank</td>
<td>37</td>
<td>66.97</td>
<td>27.11</td>
<td>15.58</td>
<td>2.56</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-50%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Posttest</td>
<td></td>
<td>94.08</td>
<td></td>
<td>8.85</td>
<td>1.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A.C.T. score in English, twenty and above--.002, below twenty--beyond .001; age, twenty and above--.290, below twenty--beyond .001; bilingualism--non-bilingualism, English only--beyond .001, English plus other--.047; attendance in The University of Arizona Model Freshman Composition Program, missed five days or less--beyond .001, missed more than five days--.002; economic rank, $8,000 or above--beyond .001, below $8,000--beyond .001; and graduation rank, fiftieth percentile or above--beyond .001, below fiftieth percentile--beyond .001.

Due to the highly significant two-tailed probabilities resulting from the t-test of the differences between the pretest and posttest means in conjunction with the eight student variables, the null hypothesis, $H_2$, was rejected. There was a significant difference between the pretest and the posttest based upon the eight student variables. The null hypothesis, $H_2$, read as follows: there will be no difference among the small groups of subjects which were compartmentalized on the basis of sex, age, grade point average in high school English, high school graduating class rank, bilingualism--non-bilingualism, family income, A.C.T. score in English, and attendance record in The University of Arizona Model Freshman Composition Program in terms of the pretest and posttest themes.
One sub-group, age above twenty years, fell below an acceptable two-tailed probability level with a computation of .290; however, the number of cases in that particular sub-group was a total of five, or only ten percent of n. Due to the very low number of cases, a lack of significance could be expected.

Results of the Step-Wise Multiple Linear Regression Analysis of the Variance on the Posttest Dependent Variable

None of the eight student identification independent variables accounted for more than a comparatively small percentage of the total variance between the pretest and the posttest. This fact is clearly illustrated in Table 6, which is a summary of the step-wise multiple linear regression test. The percentages of variance change on the dependent variable, or posttest, was accounted for as follows: pretest, 47.52 percent; attendance in The University of Arizona Model Freshman Composition Program, 5.49 percent; bilingualism--non-bilingualism, 1.55 percent; high school graduation rank, 1.55 percent; economic rank, 0.99 percent; A.C.T. score in English, 0.49 percent; sex, 0.32 percent; grade point average in high school English, 0.31 percent. The independent variable age was not computed and was automatically dropped in the course of computer analysis due to its lack of variance account and its exceedingly low tolerance level.
Table 6. Results of the Step-Wise Multiple Linear Regression Analysis with the Posttest as Dependent Variable.

<table>
<thead>
<tr>
<th>Step</th>
<th>Independent Variable</th>
<th>Multiple r</th>
<th>r Square</th>
<th>r Square Change</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pretest</td>
<td>.6894</td>
<td>.4752</td>
<td>.4752</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>Attendance</td>
<td>.7281</td>
<td>.5302</td>
<td>.0549</td>
<td>.023</td>
</tr>
<tr>
<td>3</td>
<td>Bilingualism—non-bilingualism</td>
<td>.7387</td>
<td>.5456</td>
<td>.0155</td>
<td>.217</td>
</tr>
<tr>
<td>4</td>
<td>H.S. grad. rank</td>
<td>.7491</td>
<td>.5611</td>
<td>.0155</td>
<td>.214</td>
</tr>
<tr>
<td>5</td>
<td>Economic rank</td>
<td>.7556</td>
<td>.5709</td>
<td>.0099</td>
<td>.320</td>
</tr>
<tr>
<td>6</td>
<td>A.C.T. English</td>
<td>.7589</td>
<td>.5759</td>
<td>.0049</td>
<td>.483</td>
</tr>
<tr>
<td>7</td>
<td>Sex</td>
<td>.7610</td>
<td>.5791</td>
<td>.0032</td>
<td>.576</td>
</tr>
<tr>
<td>8</td>
<td>G.P.A. English</td>
<td>.7630</td>
<td>.5821</td>
<td>.0031</td>
<td>.587</td>
</tr>
<tr>
<td>9</td>
<td>Age*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Did not compute due to inadequate tolerance level.
The established levels of significance for the step-wise multiple linear regression test were as follows: pre-test, beyond .001; attendance, .023; bilingualism--non-bilingualism, .217; high school graduation rank, .214; economic rank, .320; A.C.T. in English, .483; sex, .576; and grade point average in high school English, .587. The independent variable age, as stated above, did not compute.

The above eight student identification independent variables appeared in the order of the step-wise linear regression selection according to variance accountability.

Table 7 illustrates the arrangement of the eight student data independent variables as designed to function in the multiple linear regression analysis presented above in Table 6. Unlike the two-level presentation of the same data (Table 4) for the correlated t-test of the difference between pretest and posttest means, the variables contain both two and three level applications, as depicted by the table. Only sex, bilingualism--non-bilingualism, and attendance remain with two levels.

**Reader Reliability Levels for the Pretest and Posttest**

A Pearson product moment correlation, executed exactly as accomplished for the pre-study reader reliability coefficient test in Chapter III, revealed a reader reliability coefficient of .73 for the pretest and a reader
Table 7. Delineation of Student Data Variables for the Step-Wise Linear Regression Technique.

<table>
<thead>
<tr>
<th>Description</th>
<th>Discrimination</th>
<th>Number of Students In Group</th>
<th>Percent of n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>male</td>
<td>33</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>G.P.A. in h.s.</td>
<td>high (A,B)</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td>English</td>
<td>medium (C)</td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>low (D,F)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>A.C.T. in</td>
<td>high 20+</td>
<td>11</td>
<td>22</td>
</tr>
<tr>
<td>English</td>
<td>medium 19-15*</td>
<td>27</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>low below 15*</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Age</td>
<td>above 21*</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>18-21*</td>
<td>45</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>below 18*</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Bilingualism--</td>
<td>English only</td>
<td>40</td>
<td>80</td>
</tr>
<tr>
<td>non-bilingualism</td>
<td>English and other</td>
<td>10</td>
<td>20</td>
</tr>
<tr>
<td>Attendance</td>
<td>missed 5 or less</td>
<td>33</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>missed more than 5</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td>Economic rank</td>
<td>below $5000</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>above $5000 but below $8000</td>
<td>22</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>above $8000 but below $15000</td>
<td>17</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>over $15000</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>H.S. graduation rank</td>
<td>upper 25%</td>
<td>13</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>middle 50%</td>
<td>32</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>lower 25%</td>
<td>5</td>
<td>10</td>
</tr>
</tbody>
</table>

*Presented as two-digit figures representing actual data in numerical analyses

n=50
Table 8. Pearson Reader Reliability Correlation Coefficient for Pretest Themes.

<table>
<thead>
<tr>
<th>Variable Pair</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>.7344</td>
</tr>
<tr>
<td>AC</td>
<td>.8194</td>
</tr>
<tr>
<td>AD</td>
<td>.7061</td>
</tr>
<tr>
<td>AE</td>
<td>.5299</td>
</tr>
<tr>
<td>BC</td>
<td>.7357</td>
</tr>
<tr>
<td>BD</td>
<td>.8554</td>
</tr>
<tr>
<td>BE</td>
<td>.7456</td>
</tr>
<tr>
<td>CD</td>
<td>.7101</td>
</tr>
<tr>
<td>CE</td>
<td>.6023</td>
</tr>
<tr>
<td>DE</td>
<td>.7433</td>
</tr>
</tbody>
</table>

Reader reliability coefficient: .73
Table 9. Pearson Reader Reliability Correlation Coefficient for Posttest Themes.

<table>
<thead>
<tr>
<th>Variable Pair</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>.3461</td>
</tr>
<tr>
<td>AC</td>
<td>.4516</td>
</tr>
<tr>
<td>AD</td>
<td>.4065</td>
</tr>
<tr>
<td>AE</td>
<td>.1247</td>
</tr>
<tr>
<td>BC</td>
<td>.4238</td>
</tr>
<tr>
<td>BD</td>
<td>.4673</td>
</tr>
<tr>
<td>BE</td>
<td>.5709</td>
</tr>
<tr>
<td>CD</td>
<td>.5080</td>
</tr>
<tr>
<td>CE</td>
<td>.5056</td>
</tr>
<tr>
<td>DE</td>
<td>.4168</td>
</tr>
</tbody>
</table>

Reader reliability coefficient: .43
reliability coefficient of .43 for the posttest, as displayed in Table 8 and Table 9.

A possible solution for the apparent enigma concerning the low reader reliability coefficient level for the posttest may have been the limitation of instrument range for posttest themes. Because the papers increased in ascribed points, there was no longer dispersion from one to five on the scale. Instead, the range consisted mainly of the top three numbers, effectively eliminating the lower level of the Likert-type five-point rating instrument.

Summary

The basic findings of this chapter emanated primarily from five data analyses. These five were: (1) a numerical computation of raw score advances between the pretest and posttest Grader-assigned points on the five study criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence, (2) a correlated t-test of two means executed in order to determine the significance of statistical relationship between the pretest and posttest, (3) a correlated t-test of the difference between two means for the purpose of establishing the statistical level of significance between the pretest and posttest means of the eight student data independent variables: sex, grade point average in high school English, A.C.T.
score in English, age, bilingualism--non-bilingualism, attendance in The University of Arizona Model Freshman Composition Program, economic rank, and high school graduation rank, (4) a step-wise multiple linear regression analysis of the variance based on the posttest functioning as a dependent variable and the eight student data (above) plus the pretest functioning as independent variables, and (5) a Pearson product moment correlation analysis establishing reader reliability coefficient for the five Graders as they functioned on the pretest and the posttest using the test instrument.

The results of these analyses were as follows—(1) the numerical compilation of raw score data disclosed an increase in Grader-awarded scale points on each of the five study criteria in the following percentages: organization—thirty-five percent increase from $0_1$ to $0_2$, mechanics—sixty-one percent increase from $0_1$ to $0_2$, vocabulary—five percent increase from $0_1$ to $0_2$, communication of idea—forty-one percent increase from $0_1$ to $0_2$, logical developmental sequence—thirty-one percent increase from $0_1$ to $0_2$, (2) the correlated t-test of the difference between two means resulted in a statistically significant level of two-tailed probability beyond the .001 range for the pretest-posttest relationship, suggesting a powerful test probability for the multiple linear regression design to follow, (3) the
correlated t-test of the differences between two means for the eight student data variables and their pretest-posttest means resulted in the following levels of statistical probability for each datum—sex, male beyond .001, sex, female .029; G.P.A., A, B beyond .001, G.P.A., C, D, F .001; A.C.T., 20+ .002, A.C.T., -20 beyond .001; age, 20+ .290, age, 19-beyond .001; English only beyond .001, English plus other .047; attendance, missed 5 or less beyond .001, attendance, more than 5 missed .002; income $8,000+ beyond .001, income below $8,000 beyond .001; graduation rank above 50 percent beyond .001, graduation rank below 50 percent beyond .001, (4) the step-wise multiple linear regression test selected the eight student data independent variables (described above) and the pretest functioning as an independent variable in the following order based upon their levels of accountability for the variance on the dependent variable posttest. This data is presented in Table 10.

The pretest accounted for 47.52 percent of the variance on the posttest; the attendance factor accounted for 5.49 percent; the bilingualism—non-bilingualism factor accounted for 1.55 percent; the high school graduation rank factor accounted for 1.55 percent; the economic rank factor accounted for 0.99 percent; the A.C.T. in English factor accounted for 0.49 percent; the sex factor accounted for 0.32 percent; and the grade point average in high school English
Table 10. Variance Accountability on the Posttest Dependent Variable.

<table>
<thead>
<tr>
<th>Step</th>
<th>Independent Variable</th>
<th>$r$ Square Change</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pretest</td>
<td>.4752</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>Attendance</td>
<td>.0549</td>
<td>.023</td>
</tr>
<tr>
<td>3</td>
<td>Bilingualism--non-bilingualism</td>
<td>.0155</td>
<td>.217</td>
</tr>
<tr>
<td>4</td>
<td>H.S. graduation rank</td>
<td>.0155</td>
<td>.214</td>
</tr>
<tr>
<td>5</td>
<td>Economic rank</td>
<td>.0099</td>
<td>.320</td>
</tr>
<tr>
<td>6</td>
<td>A.C.T. English</td>
<td>.0049</td>
<td>.483</td>
</tr>
<tr>
<td>7</td>
<td>Sex</td>
<td>.0032</td>
<td>.576</td>
</tr>
<tr>
<td>8</td>
<td>G.P.A. in English</td>
<td>.0031</td>
<td>.587</td>
</tr>
<tr>
<td>9</td>
<td>Age*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Did not compute due to inadequate tolerance level.
factor accounted for 0.31 percent. The age factor was not computed due to insufficient reaction upon the dependent variable and was automatically dropped from the data during computer analysis. (5) The Pearson product moment correlation analysis established a reader reliability of .73 for the five Graders as they functioned on the pretest, dropping to .43 for the posttest. The significant decrease in reader reliability may have been due to the limitation of the range on the five-point Likert-type rating instrument from the pretest to the posttest. The Graders selected the upper end of the scale more frequently on the posttest theme ratings, thus effectively reducing its range to only the upper limits of the study instrument.

The null hypothesis \( H_1 \) was rejected due to the test results established by the percentage increase on the five study criteria. The null hypothesis \( H_2 \) was rejected due to the test results established by the t-test for the differences between \( O_1 \) and \( O_2 \) themes in conjunction with the eight student variables, with the exception of age above 20 years, which fell to .290.
CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter contains a summary of the study, a brief discussion of conclusions derived from the analyses of collected data, and recommendations for future applications and research.

Summary

This study investigated The University of Arizona Model Freshman Composition Program in specific terms with a pretest-posttest one-group design. Five Graders evaluated the $0_1$ and $0_2$ themes by means of a test instrument based on five study criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence, each criterion rated on a five-point Likert-type scale. Eight student data variables were tested for effect. These were: sex, grade point average in high school English, A.C.T. score in English, age, bilingualism--non-bilingualism, attendance in The University of Arizona Model Freshman Composition Program, family income level, and high school graduation rank.

The study data were gathered by four means: (1) the administration of pretest and posttest fifty minute
in-class writing assignments, the first sample composed early in the semester, the second late in the semester, (2) a questionnaire designed to gather three data, age, language(s) spoken in the home, and family income level, (3) a search of the records on file with The University of Arizona Registrar in order to gather grade point average in high school English, A.C.T. score in English, and graduation rank, and (4) a scanning of the class record books for the purpose of obtaining the sex of the pupil and his attendance record in The University of Arizona Model Freshman Composition Program.

The sample consisted of fifty students from three classes randomly selected out of twenty-two sections of Freshman Composition at The University of Arizona during the second semester session of the 1973-74 academic year.

A pre-study set of ten themes, typed faithfully from the original manuscripts, was rated by the five Graders using the test instrument. The result of this pre-study was the establishment of a reader reliability coefficient of .61, sufficient for continuance of the study considering the complexity of the multipartite rating requirement of the instrument.

Following the administration of the second in-class fifty minute theme assignment, all student papers were typed from their handwritten original forms, duplicating each
aspect of the writing exactly, then completely mixed so that no possible means of identifying an \( O_1 \) from an \( O_2 \) paper existed. After the mixing, the papers were transferred to the five Graders for the appropriate evaluation based on the test instrument.

Two hypotheses gave direction to the study. These were: \( H_1 \), there will be no difference between the pretest and posttest themes in terms of the criteria, and \( H_2 \), there will be no difference among the small groups of subjects which were compartmentalized on the basis of sex, age, grade point average in high school English, high school graduating class rank, bilingualism--non-bilingualism, family income, A.C.T. score in English, and attendance record in The University of Arizona Model Freshman Composition Program in terms of the pretest and posttest themes.

The data were analyzed by the following five measures: (1) a simple compilation of raw data to determine percentage increases in Grader-awarded points on each of the five study criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence; (2) a correlated t-test of the difference between two means in order to establish the statistical level of significance between the pretest and the posttest; (3) a correlated t-test of the differences between two means in order to establish the statistical level of significance between the pretest
and the posttest concerning each of the eight student variables: sex, grade point average in high school English, A.C.T. score in English, age, bilingualism--non-bilingualism, attendance in The University of Arizona Model Freshman Composition Program, economic rank, and high school graduation rank; (4) a step-wise multiple linear regression technique for the purpose of accounting for the variance on the posttest in relation to each of the eight student variables (above) and the pretest; and (5) a Pearson product moment correlation measure to establish reader reliability on the pretest and on the posttest.

The results of the five analyses were: (1) a thirty-five percent increase on the organization criterion from $0_1$ to $0_2$, a sixty-one percent increase on the mechanics criterion from $0_1$ to $0_2$, a five percent increase on the vocabulary criterion from $0_1$ to $0_2$, a forty-one percent increase on the communication of idea criterion from $0_1$ to $0_2$, and a thirty-two percent increase on the logical developmental sequence criterion from $0_1$ to $0_2$, (2) a two-tailed probability level beyond the .001 range for the pretest-posttest relationship, suggesting a powerful test possibility for multiple linear regression, (3) an establishment of a significant two-tailed probability on each of the eight student data independent variables, except age above twenty years, which fell to the .290 level, (4) an accounting of 5.59
percent of the variance between the pretest and the posttest based on the attendance variable, the remaining eight student variables accounting for less than 1.54 percent each, the most significant prediction of the posttest dependent variable being the pretest, with 47.52 percent of the variance, and, (5) a reader reliability coefficient for the pretest of .73, dropping to .43 for the posttest.

The null hypothesis, $H_1$, was rejected due to the obvious advances from the pretest to the posttest in Grader-assigned raw scale points on each of the five study criteria. A difference between the pretest and the posttest based on the five study criterion was established.

The null hypothesis, $H_2$, was rejected on the basis of results established on the t-test for the differences between the two means for each of the eight student variables. Each level of data achieved a two-tailed probability of .029 or higher, with the exception of age above twenty years, which fell to .290, possibly due to the very small number of subjects assigned to that category, which was five cases, or ten percent of $n$.

The accountable variance upon the posttest dependent variable was attributed in the main to the pretest as best predictor, with small percentages assigned to each of the eight student data independent variables.
An unnamed independent variable must be given at least a modicum of attention at this point. Although the design of the study did not permit a specific inclusion of, nor testing of, the actual planned instruction aspect of the program, that factor must be recognized as probably accounting for much of the increased Grader-awarded scale points between the $O_1$ and $O_2$ themes.

Further, the significant advances based upon each of the eight student data variables, with the exception of level one on the age factor, that being above twenty years and representing only ten percent of $n$, can be considered as associated with the instructional process extant in The University of Arizona Model Freshman Composition Program.

Conclusions

Conclusions emanating from the study were: (1) the study presented a possible model for basic investigation of a composition program, (2) the five graders, employing the five study criteria: organization, mechanics, vocabulary, communication of idea, and logical developmental sequence, unanimously awarded increased raw score points on the $O_2$ essays over the $O_1$ essays, in spite of the fact that physical recognition had been rendered impossible due to typing and thorough mixing of all papers into one homogeneous group, (3) in the case of this particular study the pretest was the
best predictor of the posttest, (4) the low reader reliability established for the posttest, .43, may have been due to the decreased latitude of the test instrument on the 0 essay, (5) a body of graders may be able to function as a coordinated unit for evaluating composition skills when parameters such as a scale and criteria have been established, and (6) the testing method may not have been sensitive enough to allow prediction concerning the variance on the posttest in relation to the eight student data independent variables; however, advances per each variable, except age beyond twenty years, was established as significant at .029 and beyond.

Recommendations for Future Applications and Research

The following recommendations are based upon the general findings of the study: (1) the study design should be employed as a model by those professional persons who would evaluate their composition programs where the use of controlled paradigms is not appropriate or practical, (2) should the study be duplicated, a scale of greater latitude should be applied, possibly one of ten points, (3) as suggested by the search of the literature, a set of criteria should be included in a composition program where the aim of the instructor is to present a clear and defined approach to improving writing skills, (4) a rating scale should be
included in a composition instruction program when the instructor is concerned with allowing the pupil an opportunity to clearly see the present level of his composition skill, and (5) the use of criteria and a rating scale should be instituted when the instructor of a composition program intends to evaluate student progress in a manner more objective than the single summary mark evaluation.
APPENDIX A

STUDENT INFORMATION QUESTIONNAIRE

DIRECTIONS: Please circle the appropriate response below:

1. Please circle the age bracket you are in.
   a. 17 years of age
   b. 18 years of age
   c. 19 years of age
   d. 20 years of age
   e. 21 or over years of age

2. Please circle the correct response.
   a. Only English is spoken in my home
   b. Only Spanish is spoken in my home
   c. Spanish and English are spoken in my home
   d. English and a foreign language other than Spanish are spoken in my home

3. Please circle the correct response.
   Our family income is:
   a. Below $5000 annual income
   b. Above $5000, but less than $8000 annual income
   c. Above $8000, but under $15,000 annual income
   d. A family income of over $15,000 annual income

Please print your name below. All information will be held in strict confidence.

Name

91
APPENDIX B

EVALUATION SHEET

Instructions: The grader is to circle the appropriate number on each of the five criteria below. The value of each number is as follows:

1. Exceedingly Inadequate (simply does not accomplish the task)
2. Inadequate (does not accomplish the task, but shows traces of improvement over grade 1 above)
3. Partially Adequate (better than 1 and 2, but still not accomplishing the goal of the criterion)
4. Adequate (not excitingly superior, but the criterion is accomplished)
5. Very Adequate (fulfillment of the criterion is accomplished)

FIVE CRITERIA (circle the appropriate number on each)

<table>
<thead>
<tr>
<th>CRITERION</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORGANIZATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MECHANICS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOCABULARY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMMUNICATION OF IDEA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGICAL DEVELOPMENTAL SEQUENCE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

CRITERIA DEFINITIONS

Introduction: The primary goal of The University of Arizona Model Freshman Composition Program is to improve expository writing. Thus, five basic criteria are employed in this study to act as a guide to clarify areas of improvement per individual student. Below is a brief commentary on the criteria, given so that an understanding of the researcher's meaning may be standardized among the various graders.

1. Organization. Crosby and Estey, College Writing (1968, pp. 55-56) find organization important from two standpoints. One, they say that "... we cannot throw all the ideas at once at our reader ... on some basis or other, we have to decide the order of the words." Second, the authors feel that because "the nature of the mind," as established by psychologists, ordering material in a logical manner aids the reader to understand and remember data. Hence, for the purpose of this study, organization will be simply the condition of order in a theme, which is conducive to logical interpretation of the writer's idea.

2. Vocabulary. Again, for the purpose of this study, a very simple definition shall be used for this criterion. Merely, it shall be considered a wise use of vocabulary when the words which are selected by the writer convey the idea with force and lively conviction.

3. Mechanics. Because of the many possible connotations of the word "grammar," it was not used in this list of criteria. However, mechanics may be construed as the following: syntax, usage, spelling, punctuation, and inflection, the nuts and bolts of the theme exclusive of idea and communication, but not without force in conveying idea and in communicating.

4. Communication of Idea. Simply, in a theme there should be one main thrust. It shall be considered the obligation of the writer to convey that idea clearly and with conviction to the audience. Difficulty in discovering what the student meant by his theme would constitute a failure in communication. Conversely, clear conveyance of idea would be success in communication.
5. **Logical Developmental Sequence.** Bacon said that "writing maketh an exact man," and it is this notion which prompts the use of criterion five. The paper should exhibit a developmental sequence which is logical. This is not organization, for it includes more. A sound, well thought out idea, developed until a point of conviction and convincing has been reached in a clear precise manner shall be considered a logical development, and if it shows a plan which builds into a solid structure, then it shall be considered a sequential development.
APPENDIX D

TOPIC SELECTION QUESTIONNAIRE

Directions: Please list below THREE topics you consider important to our times. These topics should be those which you feel will remain important for at least a year. They may be of local, national, or international importance.

1.

2.

3.
APPENDIX E

PRETEST ASSIGNMENT 01

In-Class Theme Assignment

Directions:
1. Select one of the topics listed below and write a theme on the subject.
2. The theme is to be completed in 50 minutes.
3. Place a title on your theme.
4. Be sure to put your name on your theme.

Topics:
1. The energy crisis
2. The drug problem
3. American social structure
4. Ecology problems
5. Religion in America today
APPENDIX F

POSTTEST ASSIGNMENT 02

In-Class Theme Assignment

Directions:
1. Select one of the topics listed below and write a theme on the subject.
2. The theme is to be completed in 50 minutes.
3. Place a title on your theme.
4. Be sure to put your name on your theme.

Topics:
1. Government corruption
2. Impeachment
3. Abortion legality
4. Present value of a college education
5. Generation gap
LIST OF REFERENCES


Briggs, Thomas H. "English Composition Scales in Use," *Teachers College Record*, 23 (November 1922), 423-452.


Christensen, Francis. "A Generative Rhetoric of the Sentence," *College Composition and Communication*, 16 (October 1965), 144-156.


Dusel, William J. "Planning the Program in Writing," *English Journal*, 45 (September 1956), 327.


Guffin, Jan A. "Improving Communication and Evaluation Through the Class Profile," Arizona Education Bulletin, 16 (February 1974), 142.


Herzberg, Max J. "It's No Fun to Write," English Journal, 41 (March 1952), 127-131.


Larson, Russ. "Teaching Writing to the New Students of the '70s and '80s," *Arizona English Bulletin*, 16 (February 1974), 77-82.


Sheridan, Marion C. "Can We Teach Our Students to Write," *English Journal*, 40 (June 1951), 320-324.


Thorndike, Edward L. "A Scale for Merit in English Writing by Young People," *Journal of Educational Psychology*, 2 (September 1911), 361-368.


