THE EFFECTS OF TOPIC FAMILIARITY ON THE WRITING PERFORMANCE OF NON-NATIVE WRITERS OF ENGLISH AT THE GRADUATE LEVEL

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Diane Jane Tedick, B.A., M.A.

* * * * *

The Ohio State University 1988

Dissertation Committee:

Dr. Elizabeth B. Bernhardt
Dr. Gilbert A. Jarvis
Dr. Robert N. Kantor

Approved by

[Signature]

Adviser
College of Education
Copyright by
Diane Jane Tedick
1988
To the memory of my grandparents,
Adelbert M. Mc Nerney and Vera J. Mc Nerney,
and Nana, Alice E. Osborne, who,
in their unique ways, instilled in me
the desire to learn.
ACKNOWLEDGEMENTS

Sir Isaac Newton once said that if a person accomplishes something of value it is because she stands on the shoulders of giants who have come before her. I am indebted to the cognitive psychologists and reading, writing, and language researchers whose work has influenced my views and has guided me through these pages. In addition, I am grateful to the many other giants who have touched my life. Their professional efforts and talents and their personal support were responsible for the completion of this project.

Words cannot express the respect, admiration, and gratitude I have for my adviser, Elizabeth B. Bernhardt. I say with all honesty that had it not been for Dr. Bernhardt, my life would not have followed the path that it has followed for the past four years. Dr. Bernhardt first encouraged me to pursue a degree in Foreign Language Education and throughout the past four years she has been my mentor and my friend. Her unparalleled research talents and unlimited resourcefulness have been a source of inspiration to me. As I have worked on this project, Dr. Bernhardt has guided me with her expertise, constructive criticism, and encouragement. I have been truly fortunate to have had the privilege to work with and learn from her.

Paramount levels of appreciation and acclaim are also due to the members of my reading committee, Gilbert A. Jarvis and Robert N.
Kantor. I am grateful to Dr. Jarvis for his patience and for having the insight in the early stages of my research to encourage me to examine a variety of writing measures. The study became much more interesting and enlightening as a result of his suggestion. I am also appreciative of the time he took to comment on some stylistic aspects of this work. To Dr. Kantor I owe a great deal. In 1985 Dr. Kantor hired me to teach ESL composition courses. This opportunity provided me with the experience necessary to formulate the basis for this research. I am also indebted to Dr. Kantor for allowing me to utilize the human and material resources of ESL Programs and for obtaining research funds for this project. His assistance and suggestions during the data collection were invaluable.

Many other people also offered their services during the data collection. I am grateful to the international students who participated in the study. In addition, I appreciate the efforts of all of the ESL composition instructors whose cooperation facilitated the collection of the writing samples. I also wish to thank those instructors from ESL Programs and the English Department at The Ohio State University who acted as the raters in the holistic scoring session. It was a pleasure to work with people who were so willing to give of their time and effort.

Special appreciation is extended to Diane Belcher and Julie Taaffe for the innumerable number of hours they spent doing the tedious, yet essential, scoring of each essay. Their patience, cooperation, and willingness to share their expertise and to sacrifice their time are rare qualities in colleagues. It was a privilege to work with them.
It was also a privilege to be able to conduct my research at an institution like Ohio State. The library system and computer facilities for statistical analysis are superb. I appreciate the help I received from the personnel at both the libraries and computer centers.

Many friends have offered a great deal of support throughout this endeavor. I would like to acknowledge Mike Everson for his insightful comments concerning my data analysis and for his encouragement. I also appreciate the encouragement I received from Lurdes Godinho, JoAnne Hammadou, and Miko Harada. Special thanks to Lee Cohen for asking the right questions at a particularly difficult time.

I am particularly grateful to my friend and colleague, Maureen Mathison. Maureen's insights about language and research helped to conceptualize this study. In addition, her suggestions about published works and her willingness to share her library and expertise were invaluable. Most importantly, Maureen has offered her support as a friend. Her sense of humor has helped me to keep things in perspective, and her belief in me as a teacher, researcher, and friend has helped to sustain me.

Finally, I would like to thank my parents, Howard and Vera Tedick. Their example has taught me a sense of responsibility and their faith in me has encouraged me to achieve. I am grateful for the love, solicititude, and generosity they have constantly lavished on me. Their strength and understanding are a source of pride and wonder to me.
VITA

April 20, 1956....................... Born - Cleveland, Ohio

1978................................. B.A., Kent State University, Kent, Ohio

1978-1979............................ Teacher of English as a Foreign Language, Madrid, Spain

1979-1980............................ Bilingual Itinerant Teacher (K-12), Lorain City Schools, Lorain, Ohio

1980-1981............................ Teacher of English as a Foreign Language, Madrid, Spain

1982................................. M.A., The Ohio State University, Columbus, Ohio

1982-1984............................ Bilingual Teacher (grades 3-4), Milwaukee Public Schools, Milwaukee, Wisconsin

1984-1988............................ Graduate Teaching Associate, Department of Educational Theory & Practice, ESL Programs, Department of Educational Studies, The Ohio State University, Columbus, Ohio

PUBLICATIONS


vi
FIELDS OF STUDY

Major Field: Foreign Language Education


# TABLE OF CONTENTS

- DEDICATION ......................................................................... ii
- ACKNOWLEDGEMENTS .................................................. iii
- VITA .................................................................................. vi
- LIST OF TABLES ................................................................. xii
- LIST OF FIGURES .............................................................. xiv

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. THE PROBLEM</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Statement of the Problem</td>
<td>3</td>
</tr>
<tr>
<td>Theoretical Bases</td>
<td>5</td>
</tr>
<tr>
<td>Knowledge from Cognitive Psychology</td>
<td>5</td>
</tr>
<tr>
<td>Knowledge from L1 Reading Research</td>
<td>6</td>
</tr>
<tr>
<td>Knowledge from L2 Reading Research</td>
<td>7</td>
</tr>
<tr>
<td>Knowledge from L1 and L2 Writing Research</td>
<td>8</td>
</tr>
<tr>
<td>Current Views on the Construct of Writing Proficiency</td>
<td>10</td>
</tr>
<tr>
<td>The Writer</td>
<td>10</td>
</tr>
<tr>
<td>The Reader</td>
<td>12</td>
</tr>
<tr>
<td>The Purpose</td>
<td>13</td>
</tr>
<tr>
<td>The Subject Matter</td>
<td>14</td>
</tr>
<tr>
<td>Summary</td>
<td>14</td>
</tr>
<tr>
<td>Current Views on the Assessment of Writing Proficiency</td>
<td>15</td>
</tr>
<tr>
<td>L1 Research</td>
<td>15</td>
</tr>
<tr>
<td>L2 Research</td>
<td>17</td>
</tr>
<tr>
<td>The ACTFL Provisional Proficiency Guidelines</td>
<td>18</td>
</tr>
<tr>
<td>The Test of Written English</td>
<td>19</td>
</tr>
<tr>
<td>Summary</td>
<td>20</td>
</tr>
<tr>
<td>Current Assessment Procedures Concerning Topic Familiarity</td>
<td>20</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>24</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>25</td>
</tr>
<tr>
<td>Assumptions</td>
<td>26</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>26</td>
</tr>
</tbody>
</table>

viii
## II. REVIEW OF THE LITERATURE

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>29</td>
</tr>
<tr>
<td>L1 Composition Research</td>
<td>30</td>
</tr>
<tr>
<td>Text-Based Approaches</td>
<td>30</td>
</tr>
<tr>
<td>L1 Composing Processes</td>
<td>31</td>
</tr>
<tr>
<td>Relationships Between Writing and Reading</td>
<td>34</td>
</tr>
<tr>
<td>Summary</td>
<td>35</td>
</tr>
<tr>
<td>L2 Composition Research</td>
<td>35</td>
</tr>
<tr>
<td>Text-Based Approaches</td>
<td>35</td>
</tr>
<tr>
<td>Contrastive Rhetoric</td>
<td>35</td>
</tr>
<tr>
<td>Error Analysis</td>
<td>39</td>
</tr>
<tr>
<td>Children's L2 Writing Development</td>
<td>40</td>
</tr>
<tr>
<td>L2 Composing Processes</td>
<td>42</td>
</tr>
<tr>
<td>Summary</td>
<td>46</td>
</tr>
<tr>
<td>Topic Variables</td>
<td>46</td>
</tr>
<tr>
<td>Suggested Criteria for Developing Topics</td>
<td>46</td>
</tr>
<tr>
<td>Research on Topic Effects</td>
<td>49</td>
</tr>
<tr>
<td>Wording</td>
<td>49</td>
</tr>
<tr>
<td>Mode of Discourse</td>
<td>50</td>
</tr>
<tr>
<td>Rhetorical Specification</td>
<td>50</td>
</tr>
<tr>
<td>Subject Matter</td>
<td>51</td>
</tr>
<tr>
<td>Summary</td>
<td>64</td>
</tr>
<tr>
<td>Conventional Methods for Scoring Compositions</td>
<td>65</td>
</tr>
<tr>
<td>Holistic Score</td>
<td>65</td>
</tr>
<tr>
<td>T-unit Measures</td>
<td>68</td>
</tr>
<tr>
<td>Overall Length</td>
<td>70</td>
</tr>
<tr>
<td>Conclusion</td>
<td>71</td>
</tr>
</tbody>
</table>

## III. DESIGN AND PROCEDURES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population and Sample Selection</td>
<td>73</td>
</tr>
<tr>
<td>Research Design and Variables</td>
<td>74</td>
</tr>
<tr>
<td>Materials</td>
<td>77</td>
</tr>
<tr>
<td>Experimental Procedures</td>
<td>78</td>
</tr>
<tr>
<td>The First Phase</td>
<td>78</td>
</tr>
<tr>
<td>The Second Phase</td>
<td>80</td>
</tr>
<tr>
<td>Data Collection</td>
<td>81</td>
</tr>
<tr>
<td>Holistic Scores</td>
<td>81</td>
</tr>
<tr>
<td>Preliminary Matters</td>
<td>81</td>
</tr>
<tr>
<td>The Calibration Session</td>
<td>83</td>
</tr>
<tr>
<td>Plan for Systematic Multiple Independent Scoring</td>
<td>84</td>
</tr>
<tr>
<td>The Scoring Session</td>
<td>85</td>
</tr>
<tr>
<td>T-unit Measures and Overall Length</td>
<td>87</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>88</td>
</tr>
<tr>
<td>Pilot Study</td>
<td>89</td>
</tr>
</tbody>
</table>
IV. RESULTS AND DISCUSSION........................................... 90

   Introduction..................................................... 90
   Results of the Multivariate Analysis of Variance........... 92
   Results of the Discriminant Analyses....................... 92
   Results of the Univariate Analyses of Variance.......... 98
   Discussion...................................................... 111

V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS............... 127

   Overview of the Study........................................... 127
   Summary of the Findings and Conclusions.................... 128
      Overview..................................................... 128
      Main Effects for Course Level.............................. 129
      Main Effects for Topic....................................... 132
   Implications for Assessment and Pedagogy................... 137
   Recommendations for Further Research...................... 139
   Limitations of the Study...................................... 142

APPENDICES..................................................................... 144

   A. Topic Sheet for Diagnostic Examination.................. 144
   B. Topic Sheet for Experimental Essay...................... 145
   C. Student Data Questionnaire.................................. 146
   D. Instructions for Administration of Diagnostic Examination... 148
   E. Sheet Provided for Instructors' Comments................ 153
   F. Information Requested on Outside of Sealed Examination..... 154
   G. Memorandum to Students....................................... 155
   H. Instructions for Administration of Experimental Essay..... 157
   I. Calibration Samples............................................ 161
      Level Six Samples.............................................. 161
      Level Five Samples............................................ 168
      Level Four Samples............................................ 173
      Level Three Samples......................................... 178
      Level Two Samples............................................. 182
      Level One Samples............................................ 186
   J. Sample Page of Scoring Sheet................................ 190

x
K. Multiple Independent Scoring Plan..........................191
L. Calibration Check..............................................192
M. Sample Essays Used for Scoring T-units.................193

LIST OF REFERENCES..................................................197
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Two-way multivariate analysis of variance for writing measures</td>
<td>93</td>
</tr>
<tr>
<td>2. Discriminant analysis on course level effects</td>
<td>95</td>
</tr>
<tr>
<td>3. Discriminant analysis on topic effects</td>
<td>97</td>
</tr>
<tr>
<td>4. Means and standard deviations of holistic score as a function of course level and prompt type</td>
<td>99</td>
</tr>
<tr>
<td>5. Analysis of variance of holistic score by course level and topic</td>
<td>99</td>
</tr>
<tr>
<td>6. Means and standard deviations of T-units as a function of course level and prompt type</td>
<td>101</td>
</tr>
<tr>
<td>7. Analysis of variance of T-units by course level and topic</td>
<td>101</td>
</tr>
<tr>
<td>8. Means and standard deviations of mean length of T-units as a function of course level and prompt type</td>
<td>103</td>
</tr>
<tr>
<td>9. Analysis of variance of mean length of T-units by course level and topic</td>
<td>103</td>
</tr>
<tr>
<td>10. Means and standard deviations of error-free T-units as a function of course level and prompt type</td>
<td>105</td>
</tr>
<tr>
<td>11. Analysis of variance of error-free T-units by course level and topic</td>
<td>105</td>
</tr>
<tr>
<td>12. Means and standard deviations of mean length of error-free T-units as a function of course level and prompt type</td>
<td>107</td>
</tr>
</tbody>
</table>
13. Analysis of variance of mean length of error-free T-units by course level and topic..............................107

14. Means and standard deviations of overall length as a function of course level and prompt type...............109

15. Analysis of variance of overall length by course level and topic.......................................................109
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>DESCRIPTION</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>One between-subjects (course level) one within-subjects (prompt type) mixed design</td>
<td>75</td>
</tr>
<tr>
<td>2.</td>
<td>Main effects of course level and topic as measured by holistic score</td>
<td>113</td>
</tr>
<tr>
<td>3.</td>
<td>Main effects of course level and topic as measured by mean number of T-units</td>
<td>115</td>
</tr>
<tr>
<td>4.</td>
<td>Main effects of course level and topic as measured by mean length of T-units</td>
<td>117</td>
</tr>
<tr>
<td>5.</td>
<td>Main effects of course level and topic as measured by mean number of error-free T-units</td>
<td>120</td>
</tr>
<tr>
<td>6.</td>
<td>Main effects of course level and topic as measured by mean length of error-free T-units</td>
<td>121</td>
</tr>
<tr>
<td>7.</td>
<td>Main effects of course level and topic as measured by overall length</td>
<td>126</td>
</tr>
</tbody>
</table>
CHAPTER I

THE PROBLEM

Introduction

Writing is currently viewed in academic circles as more than just a tool for communication. It is believed to improve the quality of thinking and to foster learning in students (Langer & Applebee, 1987). The growing appreciation for and interest in writing has led to numerous investigations of students' writing from grade school through graduate school. The results of these investigations, however, have not been encouraging. In fact, they have created a growing concern among educators and researchers over students' minimal writing ability. This concern has, in turn, resulted in an increase in attention to and interest in the assessment of writing ability.

In the past, large-scale writing assessment was achieved primarily by means of indirect measures, such as multiple choice items that require the examinee to recognize grammatically accurate sentences. Developers of assessment instruments have become increasingly aware, however, that indirect measures are not based upon the premise that the act of writing itself involves the production of written discourse. Test developers have, therefore, begun to view direct measures of writing, or actual writing samples, as more appropriate for measuring
writing performance because direct measures more closely approximate actual discourse.

A number of large-scale proficiency tests have added actual writing portions. Many tests developed by the Educational Testing Service (ETS), such as the National Teachers Examination, now require examinees to produce writing samples. ETS is also currently experimenting with writing portions to add to the Test of English as a Foreign Language (TOEFL). Writing portions have recently begun to be developed for national admission testing programs for law, medical, and management schools (White, 1985). In addition, large-scale assessments of writing at the local level have grown in popularity. In many universities, for example, incoming students are required to demonstrate their writing proficiency by responding to essay writing tasks. A growing percentage of the incoming students at universities are foreign students.

The population of international students is rapidly increasing in American universities—particularly in graduate programs because of the propensity Americans have for choosing the work force over graduate school (Gottschalk, 1985). Along with this increase in foreign students has been an increase in concern among members of the university community about the students' proficiency in English. University faculty in the field of English as a Second Language (ESL) have expressed the need for effective measurements of foreign students' writing and speaking skills for admission and placement purposes (Hale & Hinofotis, 1981). Similarly, in a survey of graduate faculty (Angelis, 1982, cited in Carlson et al., 1985), the writing deficiencies of foreign students were found to be of major concern.
Concern over the writing deficiencies of both international and American students has resulted in a recent emphasis on competency testing. This recent focus on competency testing and the growing insistence on large-scale writing assessment on the part of national and state government agencies and local school systems have, in turn, resulted in the need for reliable data concerning the many factors that affect the measurement of writing ability (Brossell, 1986). A factor that has recently begun to receive attention is the effect of topic variables on the assessment of writing. Topic variables have been found by writing researchers and educators to influence an individual's writing performance in critical ways. A clear understanding of the ways in which topics used for writing assessment affect writing performance is essential if accurate assessments of writing ability are to be made.

Statement of the Problem

Topic variables have become interesting to researchers and test developers largely as a result of the competency testing movement (Hoetker, 1982). Because important decisions, such as those regarding admission to and placement in universities, are made on the basis of proficiency tests, test developers are compelled to provide effective, fair topics that will elicit accurate measurements of writing proficiency.

In the past, topics used for large-scale assessment of writing were devised by national agencies, such as ETS. The recent increase of local assessments in school systems and universities across the nation
has led to rather unsystematic methods of developing topics. According to Ruth and Murphy (1988), "...it is rather common practice to borrow topics from other assessments, as if any topic could serve an all-purpose function in the measurement of writing" (p. 2). The purposes of different writing assessments, however, vary. Furthermore, as many researchers in the field of rhetoric and composition have discovered, individual writing performance varies depending upon the topic of the writing task and its underlying purpose and audience. It is imperative to gain a clearer understanding of the interaction between the writer and individual topic variables, such as wording and subject matter (Ruth & Murphy, 1988). In addition, if topic effects are to be well understood, theories of writing assessment must be closely aligned with theories of language and cognition (Ruth & Murphy, 1984). The theoretical union among these various disciplines, however, has yet to be achieved. In fact, a large gap exists between current writing assessment procedures and the knowledge that has emerged from cognitive psychology and first language (L1) and second language (L2) reading and writing research.

The present study is an attempt to apply the theoretical knowledge from the various disciplines mentioned above to the assessment of writing in learners of ESL. It investigates how L2 learners' writing performance is influenced by their prior knowledge of the subject matter of the writing task. Insights from cognitive psychology, schema theory, and writing research provide the bases for this study.
Theoretical Bases

Knowledge from Cognitive Psychology

Cognitive psychologists have known for some time that differences in experiences and knowledge affect the meaning that individuals construct from an event. In regard to the developmental stages of children, Bruner (1957), Piaget (1954), and Vygotsky (1962) spoke of the fundamental role of a child's experience in forming and organizing concepts and constructing events. In discussing learning theory, psychologists speak of the role of memory.

Both a constructive theory (e.g., Anderson & Ortony, 1975; Bransford, Barclay, & Franks, 1972) and a reconstructive theory (e.g., Bartlett, 1932; Rumelhart, 1977; Spiro, 1977) have been posited by theorists to explain the remembering process. The constructive theory states that "the contents of a memory trace are jointly determined by the sensory experience, the current knowledge structure of the learner, and pertinent features of the environmental context in which the event is perceived" (Royer, 1977, p. 168). Two learners who experience the same event, therefore, will have somewhat different memories of the event because their personal cognitive structures are different.

A variety of studies were conducted that contributed to the formation of the constructive theory. Sachs (1967), for example, in a study on listening comprehension and recall, found that subjects did not recall the exact sentences that were presented to them, but rather seemed to form a mental representation of the presented sentences and recalled their representation. Her idea was replicated by Bransford
and Franks (1971; Franks & Bransford, 1972) with written sentences. Their results supported Sachs' findings.

According to the reconstructive, or schema, theory, remembering involves the reconstruction of past events at the time of recall. It differs from constructive theory in that an event, in order to be remembered, is reconstructed on the basis of the current state of the memory schema. Spiro (1977) argues that schemata (i.e., cognitive structures) may or may not be modified by newly encountered information. In the latter case, "...it is possible that current schematic states at recall will differ from the preassimilation states at the time of comprehension" (Spiro, 1977, p. 139).

These same ideas have been applied to learning and memory theory in general (Anderson, 1977; Berliner & Rosenshine, 1977; Rumelhart & Ortony, 1977; among others). In addition, these ideas were brought forth earlier by Ausubel et al. (1968). In explaining Assimilation Theory, Ausubel et al. state that "new information is linked to relevant, preexisting aspects of cognitive structure, and both the newly acquired information and the preexisting structure are modified in the process" (p. 68, their emphasis).

Knowledge from L1 Reading Research

The above theories, particularly in relation to their explanation of the role of prior knowledge in remembering and learning, have been applied to research on the reading process. Dooling and Lachman (1971) and Bransford and Johnson (1972), for example, presented subjects with passages that were incomprehensible to the readers without provision of some kind of prior knowledge about the passages. They concluded that
subjects whose schemata were activated (by being provided with either titles or picture cues for the passages) prior to reading the passage, were able to recall much more than subjects whose schemata were not activated.

Another pioneer study on the relationship of prior knowledge to reading comprehension was conducted by Anderson et al. (1977). In their study, physical education and music majors were presented with ambiguous passages. The subjects understood the passages in terms of the mental framework that they had brought to the reading event. This study clearly illustrated the active role of the reader and his or her schemata in comprehending discourse. Most reading researchers now agree that the comprehension process is partially dependent upon the relationship between the information in the text and the reader’s existing cognitive structures (Anderson & Pearson, 1984; Spiro, 1980; van Dijk, 1982; among many others).

Knowledge from L2 Reading Research

Research findings on the L2 reading process are parallel to those of the L1 reading research presented above. Steffenson et al. (1979) conducted a notable cross-cultural study in which they investigated the comprehension of Americans and Indians reading contrived passages about an American wedding and an Indian wedding. The Indians recalled more about the Indian wedding and elaborated more (in their recalls) about the information than did the Americans. The Americans behaved in the same way with the passage about the American wedding. Johnson (1982) conducted a similar study with ESL readers using a passage about Halloween. She discovered that comprehension was higher among readers
who were assumed to be familiar with the topic due to the considerable amount of time they had spent in the United States.

Other studies (e.g., Bernhardt, 1984, 1986a, 1986b; Carrell, 1983, 1984; Hudson, 1982; Urquhart, 1984) have resulted in findings parallel to those reported by Johnson (1982). Most of the above studies have dealt exclusively with the cultural knowledge of the L2 learner. Mohammed and Swales (1984), however, investigated ESL subjects reading technical prose. They found that "field-familiarity is a much stronger indicator of rapid and successful processing than native-like competence in the language" (p. 206).

Bernhardt (1986c) created a list of generalizations that can be made about competent reading in a second language. The first generalization, based on many of the above studies, is that reading is contingent on the reader's familiarity with the topic of the discourse. Bernhardt (1986c) contends that "topic familiarity appears to be the most critical factor in comprehension--more critical, in fact, than text-based linguistic factors such as vocabulary, syntax, length of passage, and conceptual and inferential ability" (p. 26).

Knowledge from L1 and L2 Writing Research

The theoretical claims advanced by cognitive psychologists, and the findings presented by reading researchers have influenced the theoretical models that have emerged in the field of writing. Bereiter and Scardamalia (1987), for example, include content and discourse knowledge as central components of both their "knowledge-telling" model and their "knowledge-transforming" model. They contend that writers generate appropriate content by making use of the available knowledge
they have in their memories (Bereiter & Scardamalia, 1987). Similarly, Flower and Hayes (1981) posit a cognitive process theory of writing. They label one aspect of that process as planning and claim that it involves, among other sub-processes, the "act of generating ideas, which includes retrieving relevant information from long-term memory" (p. 372, their emphasis). Some writing researchers have focused on the prior knowledge features of these models and have conducted studies in an attempt to determine the extent to which prior knowledge affects how well an individual writes.

Langer (1984) and Cheskey (1984), in studies with high school students, found that subjects with high-prior knowledge of the topic produced essays higher in quality than those produced by subjects with low-prior knowledge. Parallel findings were reported by McCutcheon (1986), who conducted a similar study with elementary school children.

The only L2 study on the effects of prior knowledge on writing to date was conducted by Winfield and Barnes-Felfeli (1982). These researchers investigated the effects of familiar and unfamiliar cultural contexts on the writing of ESL subjects. They found that cultural familiarity among subjects resulted in higher measures of overall length, grammaticality, and mean T-unit length, although the latter two measures were not found to be statistically significant.

Writing theory and research have also led to a clearer understanding of the variables that are believed to represent the construct known as writing proficiency. Researchers in the field of discourse analysis have had a considerable amount of influence on current views on writing proficiency.
Current Views on the Construct of Writing Proficiency

Before an analysis of the variables that may affect writing performance can be conducted, an understanding of the construct that has been labeled as writing proficiency must be developed. A construct cannot be defined in concrete terms, but rather must be viewed on the basis of certain features that are believed to be representative of that construct; the construct itself is not directly observable. The construct of proficiency, often referred to as competence, can be thought of as one’s "presumed underlying ability," and only the "overt manifestation," or performance, of that ability can be observed (Savignon, 1983, p. 9). Hence, an individual’s writing ability is based upon inferences derived from his or her performance.

On the basis of individuals' writing performance, current discourse theorists (e.g., Britton, 1971, 1978; Brown & Yule, 1983; Kinneavy, 1971) have suggested that writing proficiency is primarily dependent on four factors. These factors are the writer (or, the speaker), the reader (or, the audience), the purpose, and the subject matter of any written discourse. Each factor may have differing degrees of influence on an individual's writing proficiency depending upon the situation and task at hand.

The Writer

The first factor, the writer, is obviously fundamental to a definition of writing proficiency. According to Purves and Purves (1986), a writer in a particular culture brings several kinds of knowledge to the activity of writing: (1) knowledge of the semantic, phonological-graphological, and lexico-grammatical features of the
language, (2) knowledge of text structures and their appropriate styles, (3) knowledge of the sociocultural norms for certain kinds of writing, i.e., pragmatics, and (4) knowledge of the subject matter of the discourse. "As a result of these kinds of knowledge, the activity then becomes a conscious and purposeful activity to bring a text into being" (Purves & Purves, 1986, p. 179).

Purves and Purves (1986) note that research in writing has primarily focused on the first kind of knowledge, linguistic knowledge. Moreover, this research has emphasized the lexical, orthographic, and syntactic elements of this knowledge. Purves and Purves suggest, however, that knowledge of text structure and knowledge of pragmatics are also important and deserving of empirical investigation. Although these latter two kinds of knowledge have received some attention in the field of contrastive rhetoric (e.g., Connor, 1987; Hinds, 1987; Kaplan, 1966, 1987) and in cross-cultural studies (e.g., Söter, 1985; Takala, Purves, & Buckmaster, 1982), additional research is clearly needed.

The final kind of knowledge that Purves and Purves (1986) mention, subject matter knowledge, has often been assumed by teachers and researchers to influence writing in critical ways (Cheskey & Hiebert, 1987; Langer, 1984; Winfield & Barnes-Felfeli, 1982). The nature of this influence, however, has been empirically investigated by only a limited number of L1 and L2 researchers.

The kinds of knowledge described by Purves and Purves (1986) illustrate the cognitive and social nature of the activity of writing. It is cognitive in that it requires thought and knowledge on the part of the writer, and social in that, as a manifestation of language, it
is an act of communication. Communication occurs when a message is sent and received. The reader, then, as recipient of the written message, is the second factor that influences writing proficiency.

The Reader

Researchers such as Birnbaum (1986), Brown and Yule (1983), de Beaugrande (1980), Petrosky (1982), Smith (1983), Tierney and Leys (1986), and Tierney and Pearson (1983), among many others, have described the interactive relationship between the reader and writer, and the processes of composing and comprehending. Petrosky (1982) provides a particularly meaningful description of this relationship:

When we read, we comprehend by putting together impressions of the text with our personal, cultural, and contextual models of reality. When we write, we compose by making meaning from available information, our personal knowledge, and the cultural and contextual frames we happen to find ourselves in (p. 16).

Petrosky's account illustrates the interactive relationship between the writer and the reader.

This relationship between the writer and the reader may function best when the text can be said to contain three criteria described by de Beaugrande (1980): efficiency, effectiveness, and appropriateness. Efficiency refers to the level of explicitness in the text. This criterion promotes "processing ease" (de Beaugrande, 1980, p. 21). In other words, the writer must be explicit enough in order to facilitate the reader's task, but not so explicit that the reader becomes bored. Effectiveness is defined as the intensity of impact on the reader. This second criterion leads to "processing depth" (de Beaugrande, 1980, p. 21); the text must engage the reader. The final criterion, appropriateness, is related to the match between the demands of the
communicative task and the extent to which the demands are met by the writer. In sum, "...the writer must balance the needs of the reader [and the demands of the communicative situation] with his or her own intention" (Sternglass, 1986, p. 160).

The Purpose

Closely related to both the reader and the writer of any discourse is intention, or purpose, the third factor that influences writing proficiency. Purpose is clearly an integral part of the first two factors described above. The writer's purpose—to inform, to persuade, to express—will certainly lead the writer to tap different sources of knowledge as s/he is engaged in the activity of writing. Similarly, the reader's purpose—to learn, to assess, to enjoy—will lead the reader to predict and to make decisions about the text as s/he is engaged in the activity of reading. The reader may be further influenced by the writer's purpose, particularly if the reader is in an assessment situation and must decide how well the writer accomplished his or her original intention.

Kinneavy (1971) argues that purpose is the most influential factor in guiding a writer to make choices in regard to content, organization, and form during the composing process. Both he and Lloyd-Jones (1979) suggest that "skill in accomplishing one rhetorical purpose does not necessarily imply skill in accomplishing another" (Odell et al., 1978, p. 3). Moffett (1968) and Gibson (1969), on the other hand, claim that one's sense of the interrelationship of speaker, audience, and subject matter is a greater influence on a writer's choices.
The Subject Matter

The final factor influencing writing proficiency is the subject matter of the written discourse. Again, this factor is closely related to the writer, reader, and purpose. Different subject matter leads to different underlying purposes for writing. Some subject matter requires persuasive or analytic discourse, whereas other subject matter demands informative or expressive discourse. There is no doubt that individuals' writing proficiency varies considerably according to the mental operation (i.e., analysis, description, narration of a story) required by the subject matter (White, 1985).

The extent to which the writer and the reader are familiar with and interested in the subject matter of the written discourse will affect the presumed level of proficiency of that discourse. Ruth and Murphy (1984, 1988) argue that the subject matter of a particular topic may be interpreted in different ways. In an assessment situation, a test-maker has a particular interpretation in mind when developing a topic, which may be different from the test-taker's (i.e., the writer's) interpretation of that same topic. These interpretations, in turn, may also be different from the test-rater's (i.e., the reader's) interpretation. The varied interpretations are likely to have a profound effect on the judgment of that writer's level of proficiency.

Summary

The four factors influencing writing proficiency, writer, reader, purpose, and subject matter, are interactive and interdependent. A clear understanding of the relationship of these four factors and the effects each has on writing proficiency has yet to be achieved. White
(1985) reminds researchers and educators that "...we must keep in mind that writing proficiency is a question, not an answer--a problem for definition, not a solution to the problem. We move most effectively toward the answers we need when we see the term in a context that it simultaneously reflects and changes" (p. 58). A meaning-based, albeit tentative, definition of writing proficiency that incorporates the four factors described above is offered by Odell (1981). He defines writing proficiency as "...the ability to discover what one wishes to say and to convey one's message through language, syntax, and content that are appropriate for one's audience and purpose" (p. 103). This definition is particularly relevant for the present study.

Current Views on the Assessment of Writing Proficiency

L1 Research

As mentioned earlier, test developers have recognized that direct measures (e.g., actual writing samples) are more appropriate than indirect measures (e.g., multiple choice questions) for assessing writing proficiency. Actual writing samples are believed to be more appropriate with regard to the issue of face validity. This belief in the face validity of writing assessment instruments has led test developers to direct their research efforts more toward issues of reliability than toward issues of validity.

Almost two decades ago, Coffman (1971) and McColly (1970) noted that the vast majority of researchers' attention has been devoted to issues of reliability, not validity. This predominant interest among test developers in the issue of reliability has resulted in a great
number of studies on scoring procedures. These studies have focused primarily on the raters of the writing samples and have resulted in lists of suggestions for training raters to score essays holistically (e.g., Braddock et al., 1963; Cooper, 1977; Godshalk et al., 1966; Henning, 1984; Jacobs et al., 1981; White, 1985). In addition, evidence for and against holistic scoring procedures has emerged from the literature (e.g., Braddock et al., 1963; Cronbach, 1971; Diederich et al., 1961; Follman & Anderson, 1967). Other scoring procedures, such as Primary Trait Scoring (Lloyd-Jones, 1977) and Analytic Scoring, (Spandel & Stiggins, 1980) have also been developed.

This focus on reliability has also led to studies comparing indirect measures and direct measures of writing proficiency (e.g., Coffman & Papachristou, 1955; Godshalk et al., 1966; Huddleston, 1954). The premise underlying these earlier studies was to make a case for indirect measures because of their high reliability. The primary emphasis of these studies was on the scoring, not the content, of the instruments. More recently, White and Thomas (1981) conducted an experiment to compare the results of multiple-choice and essay examinations with a different purpose in mind. Their premise was to examine the fairness of the tests with regard to ethnic minorities. They concluded that "essay testing is substantially more fair than multiple-choice testing to racial and ethnic minorities" (White, 1985, p. 72). Their conclusion is subject to criticism, however, in light of evidence of sociocultural bias in essay topics (Brick, 1980, as cited in Brown, 1986).
Recently, Hoetker (1982) and Ruth and Murphy (1988) have echoed the 20-year old statements made by Coffman (1971) and McColly (1970) that research has dealt almost exclusively with issues of reliability. Hoetker (1982) notes that the issue of validity has been almost totally ignored, as have other sources of error in essay examination results, such as the writer and the topics. Issues regarding essay prompts have tended to be neglected. In fact, Hoetker (1982) observes that "until the past few years, reports of research studies and testing programs usually did not bother to include the texts of essay topics" (p. 380). This lack of interest in topic effects on the part of test developers finally has begun to be criticized.

L2 Research

The research on the assessment of L2 writing has followed a path similar to that of L1 writing assessment. The merits of essay examinations have been recognized. Large-scale proficiency examinations are beginning to require examinees to produce actual writing samples. L2 researchers have also focused much of their attention on devising methods for reliability of scoring the writing samples. Furthermore, reliability of scoring has often been examined on the basis of a comparison between indirect and direct measures (e.g., Kaczmarek, 1980; Flahive & Snow, 1980; Pike, 1979). The issue of validity has been primarily confined to scoring procedures. Both Cooper (1977) and Perkins (1983) attest to the validity of the holistic scoring procedure, yet neither based his statement on empirical findings. In a recent study with raters of ESL compositions
(Janopulos, 1987), however, empirical data were provided to support the construct validity of the holistic scoring technique.

In sum, focus has been on the scoring of writing performance, not on the testing instruments used to elicit that performance. The following descriptions of two well-known attempts to assess L2 writing proficiency illustrate this point.

The ACTFL Provisional Proficiency Guidelines. Perhaps the most well-known recent attempt to assess L2 writing proficiency is that of the ACTFL Provisional Proficiency Guidelines (1984). These guidelines appear to be based on the belief that there exists a generic proficiency for all language modes with which an individual's proficiency can be compared. Furthermore, this proficiency can be measured on the basis of an individual's performance with regard to tasks that have presumed inherent degrees of difficulty. For example, the Novice-High level of the guidelines for writing proficiency states that "[the subject] can supply information...on forms such as hotel registrations and travel documents...[and] can write...short phrases and simple lists" (p. 223). An examinee showing a proficiency classified as Advanced-Plus can "write about most common topics...[and] write résumés...summaries [and]...most social and informal business correspondence" (p. 225).

These examples clearly illustrate the fact that the ACTFL Guidelines do not take into account the four factors (i.e., writer, reader, purpose, and subject matter) considered by discourse theorists as essential components of an individual's writing proficiency. The tasks required of the writers have not been developed on the basis of
empirical data. In fact, the writing tasks appear to have been
designed to promote ease and reliability of scoring: The tasks have
been designed for the readers, not the writers.

The Test of Written English. The second well-known attempt to
design an instrument to assess L2 writing proficiency is the Test of
Written English (TWE). The TWE, currently in experimental stages, is
the writing portion of the TOEFL. Developers of the TWE (Carlson et
al., 1985) made use of several perspectives in devising their working
definition of writing proficiency. Among these perspectives appears
functionally based communicative competence. Communicative competence,
a term coined by Dell Hymes (1971), involves the ability to use
language that is appropriate in a particular situation in order to
communicate effectively. Carlson et al. (1985) add that it is
"'functional,' in that it 'works,' serving to convey what the person
intended and resulting in appropriate receptive behavior (thought or
action) by the recipient of the communication" (p. 6). Savignon
(1983), in discussing communicative competence, describes it as
dynamic, context specific, relative, and dependent on appropriate
register and style on the part of the writer (or speaker).

The development of writing tasks for the TWE was based on a list of
criteria developed by Carlson et al. (1985). One criterion states that
the mode of discourse be specified in the prompt. Carlson et al.
(1985) based their choice of mode of discourse on the results of a
survey of academic writing tasks conducted by Bridgeman and Carlson
(1984). The topic types chosen were (1) comparison/contrast plus take
a position, and (2) describe and interpret a graph or chart. Topics
were carefully developed and selected on the basis of pretests. The list of criteria upon which the actual content of the topics was based, however, was not derived from a strong research base. Carlson et al. (1985) admitted that "much research remains to be conducted regarding to what extent the parameters of a writing assessment instrument influence writing performance" (p. 11).

Summary

It is clear from the above discussion on current assessment procedures that research on the validity of the instruments used for testing writing is sorely needed. One aspect in need of attention is how writers' familiarity with the topic of a writing task affects their performance. Although important findings on this aspect of topic variables have been reported by composition researchers (e.g., Cheskey, 1984; Langer, 1984), these findings have had little influence on the current treatment of topic familiarity in assessment situations.

Current Assessment Procedures Concerning Topic Familiarity

Variation due to unequal familiarity with the topic of the writing stimulus is not a new problem for test developers. Coffman (1971) addressed this problem almost two decades ago and cited Wiseman and Wrigley's (1958) suggestion that writers be given a choice of several topics on an examination when being asked to demonstrate their writing ability. Coffman (1971) was careful, however, to point out the difficulties that this approach poses to the reliability of scoring. Furthermore, this approach has since been investigated and has not been
found to be an effective method for controlling for topic familiarity (e.g., Evans, 1979; Freedman, 1983; White, 1985).

At approximately the same time Coffman's (1971) article appeared, McColly (1970) addressed the risks to validity produced by unequal familiarity with the topics. He criticized the use of traditional, highly structured topics (i.e., those which provide detailed description prior to the question). He argued that "...in giving all students a basic content, and asking them in effect to make something of it, the examination task becomes really a task in logic more than one in pure writing" (p. 153). The alternative McColly suggested was "a topic in which...students are deprived of something to say" (p. 153, emphasis added). Examples of the types of topics he suggested follow.

Consider these contradictory proverbs: "Look before you leap," and "He who hesitates is lost." Decide which of the two offers better advice, then write an essay in which you defend your choice.

You have heard the saying, "The best things in life are free." Decide whether this is true or false, then write an essay in which you defend your opinion (p. 153).

McColly's underlying premise was, "...for a writing-test topic to be valid, it should have the property of filtering out not only differences ascribable to knowledge, but also those arising from fluency in logical operations" (p. 153). Examples of this alternative kind of topic were tested in an experiment (McColly & Remstad, 1963, cited in McColly, 1970) and were not found to be any more valid than the highly structured topic may have been (McColly, 1970).

More recently, test developers have attempted to address the considerations in topic development that were described earlier by devising writing stimuli that meet the assumed criteria of good
topics. In recognizing the problem of unequal familiarity with
topic, test developers have made sincere attempts to make the exams
fairer by creating general topics in order to equalize the writers' familiarity with the topic. Carlson and Bridgeman (1986), for example,
state that "the content implied by the topic must be as fair as
possible, not favoring a specific set of personal or cultural experiences" (p. 139). It appears that the principle underlying test
developers' use of general topics is that the writers will have
approximately the same amount of familiarity with the general topics,
which will, therefore, equate their chances for success. An example of
an essay question used on the College Board follows.

"We must live in the present. If we dwell on the past, we will
lose the present." To what extent and in what ways do you agree
or disagree with this statement? Explain and illustrate your
answer from history, literature, observation, or experience
(Fader, 1986, p. 88).

The above topic has also been used as a prompt for an ESL diagnostic
examination in a large university program. Another topic used in the
same ESL program follows.

Many people believe that it is better to act quickly and
decisively than to wait and think something over carefully,
because you may lose the opportunity by waiting. Do you
think taking quick, decisive action is or is not advisable?
(R. Kantor, personal communication, July, 1987).

These examples and other topics often used in large-scale testing
programs bear a strong resemblance to those topics suggested by
McColly (1970). Recall, however, that McColly's premise was to
device topics about which writers would have little prior knowledge.
Test developers, then, in their attempt to provide all writers with
the same advantage, have, in fact, put all writers at a disadvantage by not allowing them to make use of their prior knowledge.

The previous statement is supported by experts in the field of measurement. In an extensive discussion of validity, Cronbach (1971) notes that in order for proficiency tests to have construct validity, the content of the tests must allow for the examinaes to demonstrate their actual ability. In reference to writing assessment, Coffman (1971) states, "...since it is the students' underlying ability to express ideas in written form one wishes to assess, it is only fair to permit him to demonstrate the ability with a subject for which he has an affinity" (p. 289). Measurement scholars in the field of language assessment voice similar assertions. Johnston (1983), for example, questions the validity of the current standardized procedures for the assessment of reading comprehension because, among other things, the instruments lack consideration of the readers' prior knowledge. Similarly, White (1985) explains that the construct validity of writing assessment instruments can be questioned because of differences in the amount or quality of relevant knowledge the examinees have available to perform effectively on a particular topic. Oller (1979), with specific reference to L2 writing tasks, states that "...such tasks will be maximally successful when the writer is motivated to write about something that has personal value to himself and that he would want to communicate to someone else..." (p. 384).

The previous statements confirm the need to question the validity of the current procedures used for the assessment of writing proficiency. The present study investigates the validity of the
current procedures used to test the writing proficiency of L2 learners. It focuses on the effects of topic familiarity on the writing performance of ESL writers at three different course levels.

Purpose of the Study

The primary purpose of the present study, stated in terms of a research question, is:

Does topic familiarity have an effect on the writing performance of nonnative writers of English who are at different course levels, as measured by indices of holistic scores, T-units, error-free T-units, mean length of T-units and error-free T-units, and overall length?

In addition to answering the major research question, the study has the following objectives:

1. To provide data to contribute to the presently limited L1 and L2 literature base on topic development for the assessment of writing;

2. To provide documentation of developmental aspects of writing ability. This will be accomplished by means of the quantitative analyses of the differences exhibited by nonnative writers of English at three different course levels as measured through the selected dependent variables mentioned above.

Significance of the Study

The significance of the present study is that it will provide data to add to the exceedingly limited amount of L1 and L2 writing research concerning the effects of topic familiarity on writing ability. Because this study focuses on large-scale writing assessment, its
results will be of interest to the growing literature base on writing assessment in general. As mentioned previously, research is clearly needed on aspects of topic variables, such as wording, mode of discourse, rhetorical specification, and subject matter. In their respective discussions about the nature of the effects of writing prompts upon performance, Brossell (1986), Hoetker (1982), Ruth and Murphy (1988), and White (1985) have recognized this need and have called for further research. Hoetker (1982) and White (1985) emphasize the need for research conducted under assessment circumstances. The methods and writing tasks used successfully in the teaching of writing or in classroom research may be totally inappropriate for the assessment of writing because of the different purposes underlying each situation. To date, few L1 studies and virtually no L2 studies have investigated topic variables under assessment circumstances.

This study will be the first large-scale investigation of ESL adult learners writing expository text about their fields of study. Until this time, the only published L2 study on topic familiarity examined the effects of culturally familiar context on written recalls.

Because this experiment examines the writing of subjects in three different course levels of ESL composition, it will provide important data to indicate at which level(s) topic familiarity makes a difference in terms of the measures to be examined. In other words, subjects at the various levels may benefit from topic familiarity in different ways.

When viewed from a pedagogical perspective, the present study can begin to answer questions concerning some of the practices currently
employed in L2 writing programs—particularly with respect to examination practices. Certain topics used to elicit writing performance, for instance, do not allow writers to make use of their prior knowledge. One of the questions this study will attempt to address is the efficacy of such a practice.

**Assumptions**

The present study is based on the following assumptions:

Subjects will cooperate by putting forth their best effort in responding to both prompts;

Subjects' written responses accurately reflect their writing ability;

Subjects' placement into the various course levels is an accurate reflection of their actual level of writing proficiency;

For the field-specific prompt, subjects will choose topics with which they are familiar;

Information obtained from the personal data questionnaire is accurate.

**Definition of Terms**

The following terms as used in the present study are defined as follows:

*Holistic Scoring* — method of assessing writing proficiency whereby trained raters quickly read compositions and score them based upon their overall impression of the quality of the essays. Scoring will be accomplished on a six-point scale, from one (reflecting the performance
of a nonnative writer having a low proficiency) to six (reflecting the performance of a nonnative writer having a high proficiency). Each level of the scale will be defined on the basis of sample essays written by nonnative writers who are not chosen as subjects for the study. 

**Calibration Session** - training session during which raters read and rate sample essays that correspond to the levels of the scale to be used for holistic scoring. The goals of this session are "to obtain agreement on the scores of sample papers and...to help the readers internalize the scoring scale..." (White, 1985, p. 25).

**T-unit** - index of syntactic maturity defined "as a single main clause...plus whatever other subordinate clauses or nonclauses are attached to, or embedded within, that one main clause" (Hunt, 1977, pp. 92-93).

**Error-free T-unit** - modification of the basic index, the T-unit, that is used as a measure of growth in L2 research. In the present study, for a T-unit to be judged as error-free, it must not contain morphological or syntactic errors, or errors related to article usage, and it must make contextual sense. Spelling and punctuation errors are permitted, as well as some lexical errors, depending upon their degree of seriousness. The lexical errors permitted are based upon the judgment of the experienced raters; a lexical error is permitted if it does not grossly distort the meaning being communicated by the writer.

**Mean Length of T-unit/Error-free T-unit** - index measured by the mean number of words per T-unit/error-free T-unit in an essay. Acronyms and
names of countries are counted as one word. Hyphenated words, such as
time-consuming, are counted as two words.

Overall Length - index of fluency as measured by the total number of
words in an essay. The same stipulations used for mean length of
T-units/error-free T-units are used for this measure.
CHAPTER II
REVIEW OF THE LITERATURE

Introduction

Topic variables are believed to have a considerable amount of influence on an individual's writing performance. The nature of this influence, however, is not clearly understood. Because writing is an activity that is dependent on the prior acquisition of knowledge, an individual's writing performance will be influenced by topic variables in different ways. For this reason, topic variables are exceedingly complex; they interact differently with different writers. As explained earlier, writers need to tap several kinds of knowledge in order to participate successfully in the activity of writing. These kinds of knowledge include knowledge of language, knowledge of discourse structure, knowledge of pragmatics, and knowledge of subject matter (Purves & Purves, 1985).

The different kinds of knowledge mentioned above interact with the factors associated with the topic that is used to elicit writing performance. These interactions lead to possibly different interpretations of the same topic by a number of writers and to the subsequent writing performance, which also varies, depending upon the
degree to which an individual writer's realm of available knowledge fits the task.

Research on topic variables is in its infancy. Much of the composition research in general has influenced the directions of research on topic variables. The general composition research has provided valuable insights about the different kinds of knowledge that writers bring to the writing task and about the ways they use that knowledge to communicate their message. Insights from both L1 and L2 composition research have influenced research on topic variables in important ways. One branch of the composition research has dealt with the written text itself and another has focused on the writers.

L1 Composition Research

Text-Based Approaches

A number of researchers have attempted to develop procedures for analyzing written texts. Halliday and Hasan (1976), for example, have examined texts according to cohesive elements. Their categorization of cohesive ties has been used to compare the writing of good and poor college writers (Witte & Faigley, 1981) and to trace the development of narrative writing in children (King & Rentel, 1981). These studies have provided insights about the influence of linguistic and discourse knowledge on writing ability.

Several researchers have devised models to describe the organization of texts (e.g., Kintsch, 1974; Meyer, 1975, 1981). These models function on the basis of the hierarchical relationships among the propositions in written discourse. The models have been used to
examine the information that readers recall from written text and, more recently, to examine the quality of compositions. The analysis of written recalls is important in that it shows how a writer makes use of his or her prior knowledge to make meaning in the act of reading and in the act of writing.

Yet another attempt to investigate text has been focused on the analysis of written errors. Shaughnessy's (1977) seminal study of the errors made by basic writers led scholars in the field of writing research to view errors positively. Errors are now seen as sources of learning for writers and as sources of information for teachers and researchers. Shaughnessy's (1977) work became influential in the movement during the 1970s to view writing not only as a means for demonstrating learning, but also as a tool for learning in and of itself (Emig, 1977). The research most characteristic of this movement is the research on composing processes.

**L1 Composing Processes**

A classic study conducted by Janet Emig in 1971 represented one of the first attempts to discover what writers do as they compose. Since that time, numerous studies on composing processes have been conducted. In addition to the high school writers examined by Emig (1971) and others, researchers have examined a variety of groups of writers: experienced, professional writers (Emig, 1975; Murray, 1978; Sommers, 1980), college writers (Flower, 1979; Flower & Hayes, 1977; Pianko, 1979, Sommers, 1980), children writers (Bissex, 1980; Britton et al., 1975; Calkins, 1983; Clay, 1975; Graves, 1975; Harste et al., 1984), unskilled writers (Perl, 1979), and teachers (Perl, 1980). These
researchers have come to the consensus that composing processes are non-linear, recursive, and interrelated. "...The tasks of planning, retrieving information, creating new ideas, and producing and revising language all interact with one another throughout composing" (Flower & Hayes, 1980, p. 32). The above studies have provided valuable information about how writers' different kinds of prior knowledge affect the act of writing.

The studies on composing processes have also led to descriptions of skilled and unskilled writers. Unskilled writers spend less time planning (Pianko, 1979), and their plans are not as flexible as those of skilled writers (Rose, 1980). Unskilled writers reread their work less frequently than skilled writers, and when they reread, they do so more for purposes of correcting surface-level errors than for concentrating on whether they are communicating effectively (Faigley & Witte, 1981; Flower & Hayes, 1981; Perl, 1979; Sommers, 1980). In addition, they are excessively and prematurely concerned with avoiding errors (Perl, 1979). They rarely rework the ideas set forth in the first draft and regard revision as little more than editing.

Several researchers (e.g., Beach, 1982; Elbow, 1981; Flower, 1979; Flower & Hayes, 1981) have added that skilled and unskilled writers attend to audience in different ways. Unskilled writers and children in early stages of their development engage in what Flower (1979) has termed "writer-based prose," i.e., text that is written with little or no consideration of audience. This kind of writing puts a considerable amount of responsibility on the reader to make meaning from the text. In contrast, "reader-based prose" (Flower, 1979) represents text
written with careful consideration of a particular audience. Flower (1979) contends that "reader-based prose" should be the writer's goal, but that "writer-based prose" is a necessary and informative stage in the process of achieving that goal. Attention to audience has been designated as important to emphasize in writing—particularly during the revision process (Beach, 1979, 1982; Elbow, 1981; Flower, 1979; Roen & Willey, 1988; Sommers, 1980). The above suggestions have led some researchers to believe that audience specification in writing tasks can affect writing performance in critical ways. This will be an important area of inquiry in research on topic variables.

Another area of inquiry that has emerged from research on composing processes has been that of response to writing. Results from studies conducted on response to writing have been less than encouraging. Researchers have found that teachers impose their ideas on students' compositions (Brannon & Knoblauch, 1982; Moran, 1981; Sommers, 1982). Furthermore, teachers apply uniform, inflexible standards to their students' compositions and respond to those compositions according to the degree to which they meet the prescribed standards (Moran, 1981). Teachers also often misread compositions, as evidenced by their frequently inaccurate and inappropriate comments (Sommers, 1982). Teachers view texts as final products; they attend to surface-level features in compositions that should be considered as first drafts (Butturff & Sommers, 1980; Moran, 1981; Murray, 1982; Sommers, 1982). This practice leads student writers to assume that surface-level features are most important, and may be one reason why unskilled writers are overly concerned with surface-level errors. In addition,
responses have been found to be contradictory, inconsistent (Sommers, 1982) and vague (Butturff & Sommers 1980). These findings are important to research on rater variables in writing assessment. Test raters and test takers may be influenced by the topic in different ways, which may, in turn, lead the rater to misjudge the writer’s performance (Ruth & Murphy, 1984, 1988).

The above researchers, who have focused on the effects of response and audience awareness on writing, have been primarily interested in the relationship between the writer and someone different from the writer, who acts as the reader. Other researchers have been interested in the relationship between the writer and the reader when they are one in the same. In other words, these researchers have focused on how individuals’ experiences with writing affect their reading, and, similarly, how individuals’ experiences with reading affect their writing.

Relationships Between Writing and Reading

Studies in this area of research provide valuable information in regard to the ways in which a writer’s prior knowledge of reading affects his or her writing. Test developers will need to consider the relationship between reading and writing in developing effective topics for the assessment of writing.

Researchers have found that the quality of reading is related to the quality of writing. This relationship has been observed in college subjects (Spivey, 1983), elementary and middle school children (Birnbaum, 1982), and in children of low socio-economic status (Chall & Jacobs, 1983). Similarly, a longitudinal, ethnographic study of first
graders (Graves & Hansen, 1983) revealed strong, consistent relationships between their reading and writing development. In a similar vein, other researchers have provided evidence that the structure of reading passages affects students' writing. Eckhoff (1983), for example, found that children's writing contained elements of the stilted language found in their basal readers. Other researchers have provided evidence that the format of texts influences the format of students' writing in high school (Geva & Tierney, 1984) and in elementary school (Gordon & Braun, 1982). Related studies have shown that writers acquire rhetorical knowledge (Bereiter & Scardamalia, 1984) and style (Church & Bereiter, 1983) from their reading.

Summary

The brief review of literature above mentions only some of the insights from L1 composition research that have influenced research on topic variables. Insights from L2 composition research, which in many ways parallel those from L2 composition research, show that topic variables are in great need of research in L2 contexts. L2 studies on writing, like L1 studies, have been focused on both the text itself and on the writers.

L2 Composition Research

Text-Based Approaches

Contrastive Rhetoric. An interesting area of inquiry that has recently begun to receive serious attention is that of contrastive rhetoric. The research on contrastive rhetoric provides evidence of
the nature of cultural experience, i.e., knowledge of pragmatics, on writing. This field began with the rather controversial publication by Kaplan in 1966. In this publication, Kaplan (1966) argued that different cultures produce different rhetorical patterns in their writing. He based his argument on compositions written by Chinese, Korean, Arabic, and European college students of ESL and on translations of documents written in other languages. Kaplan (1966) has been attacked on several counts.

First, Kaplan (1966) bases his conclusions on paragraphs, not on whole texts. Secondly, he makes broad generalizations about the different cultural groups based on relatively few samples of their writing. More importantly, the samples Kaplan used were compositions written in English—the writers' L2. They were, therefore, not indicative of the writers' L1 rhetoric. It is false to assume that if an individual writes well in his or her L1, s/he will do the same in an L2 (Mohan & Au-Yeung Lo, 1985). Furthermore, neither researchers nor teachers should assume that L2 students are competent writers in their L1. In other words, notions of transfer and interference must be dealt with cautiously. Thirdly, Kaplan classified different cultural groups together under broad umbrellas. For example, he argues that generalizations he draws about paragraphs written in Arabic "...would be more or less true for all Semitic languages..." (Kaplan, 1966, p. 46). In addition, he claims that Oriental patterns are true for Chinese and Korean writers, yet provides examples only from Korean writers in his article.
Kaplan (1987) has since responded to the above criticisms and has admitted to having "...overstated [his] case" (p. 9). Kaplan (1987) has modified some of his original claims and has reiterated others. He states now that the rhetorical patterns presented in his 1966 article are possible in all languages, depending upon the situation surrounding the creation of a particular text. In addition, Kaplan (1987) recognizes that in order to make claims about texts, it will be necessary to examine long passages, not just sentences or paragraphs. There also needs to be added focus on contrasting the composing processes of writers, not only their written products. Some of Kaplan’s (1966) earlier arguments are acceptable. Rhetorical patterns vary from culture to culture because cultures are different. Language and culture are inseparable. Part of learning a language involves mastering its system of logic, which evolves out of culture (Kaplan, 1966).

Despite the controversy surrounding Kaplan’s earlier notions of contrastive rhetoric, researchers have begun to explore contrastive studies of texts. Connor (1987) examined persuasive and argumentative texts written by subjects from England, Finland, Germany and the United States. She argues that, in order to analyze the quality of texts, it is essential to view the texts from linguistic, psycholinguistic, and sociolinguistic perspectives. Connor’s (1987) study was part of a mammoth project funded by the International Association for the Evaluation of Educational Achievement, which involved 20,000 compositions written by students from 16 countries (Purves & Takala, 1982). Another part of this same study conducted by Takala, Purves,
and Buckmaster (1982), which analyzed the large set of essays, found similarities and differences among the compositions. Takala et al. (1982) came to one general conclusion: Writing is both a social and cognitive phenomenon. Writing samples must be analyzed, then, from a variety of perspectives.

Other studies in the field of contrastive rhetoric have focused on descriptions of specific languages. Hinds (1987) discusses the different interpretations the Japanese and Americans have toward the responsibilities of the reader and the writer. In Japanese discourse, more responsibility is placed on the reader to gain meaning from text, whereas in English, the writer is faced with the obligation to make his or her meaning clear to the reader.

Eggington (1987) has explored conflicting expository style produced by Korean writers in Korea. One style is derived from traditional Korean rhetoric, and the other is influenced by American rhetoric. This latter style is common in the writing of Koreans who have studied in the United States. Eggington (1987) claims that these different patterns may hinder optimal communication between readers and writers in Korea.

Ostler (1987) examined ESL compositions written by native speakers of Arabic. She provides evidence to support Kaplan's (1966) earlier notion that Arabic speakers tend to strive for elaborate parallel structures (i.e., an equal balance between subjects and predicates).

Yet another branch of the field of contrastive rhetoric has emerged in recent years. Some researchers have argued that fields of study constitute sub-cultures of their own and that the rhetorical patterns
in these sub-cultures must be described and analyzed. Grabe (1987) makes a distinction between expository prose written in the sciences and humanities. He argues that finer distinctions among text types must be made in English as well as in a variety of other languages before an actual basis for rhetorical comparison is possible. This notion of rhetorical types with fields of study has also been addressed with regard to reading difficulties faced by L2 learners (Widdowson, 1979).

The above studies on contrastive rhetoric provide information that needs to be considered for the development of effective L2 writing assessment instruments and scoring procedures. Additional research is needed, however, before definitive solutions can be developed to address the existing assessment difficulties that are due to cultural differences.

Error Analysis. Another text-based approach to analyzing L2 written discourse is that of Error Analysis. Error Analysis in L2 research used to involve little more than identifying and linguistically classifying learners’ errors. It has since come to be used as a way of gaining information about the process of language acquisition. Ellis (1985) emphasizes that Error Analysis provides both linguistic and psycholinguistic information about the phenomenon known as interlanguage. The second type of information is helpful to researchers in understanding the sequential development that occurs in L2 acquisition. Although the majority of interlanguage studies involving Error Analysis are based on spoken discourse, some studies on written discourse have added to the interlanguage literature (e.g.,
Adjémian & Liceras, 1984; Krashen, 1978; Mazurkewich, 1984). Perhaps one of the most positive results to have emerged from the new perspective on Error Analysis is that errors are no longer viewed negatively; errors are viewed as natural and necessary for L2 acquisition and as informative for L2 researchers. Furthermore, because written errors are now viewed in a different perspective, writing is beginning to have a place in L2 classrooms, whereas in the past it was commonly postponed until learners had achieved a level of competence in speaking, listening, and reading.

The research on Error Analysis has provided researchers with valuable information regarding the developmental characteristics of writers. The extent to which developmental characteristics interact with topic variables will be an important area for future inquiry.

The realization that errors provide information about the writers' language development has also led to an increased focus on the writers themselves. A branch of L2 composition research that has focused on the writers has been the research involving L2 writing development in children.

**Children's L2 Writing Development**

Research on the development of L2 writing ability in children clearly illustrates that prior knowledge affects writing in critical ways. Many of the studies on L2 writing in children have led to the belief that what children know about writing in their L1 helps rather than hinders their L2 writing development (Edelsky, 1982, 1983; Hudelson, 1983, 1984a, 1984b, 1984c; Saville-Troike, 1984a, 1984b). Edelsky (1982, 1983) examined the L1 and L2 writing development of
children in a bilingual (Spanish/English) program by analyzing a number of text features found in over 400 compositions. These features included segmentation and punctuation, code-switching, spelling inventions, and structural and content features. Edelsky (1983) found that bilingual children segment words on syntactic and phonological/morphological bases in both their L1 and L2 writing. L2 learners also demonstrate unconventional patterns of punctuation and, like L1 learners, overgeneralize in the use of punctuation (Edelsky, 1983). Hudelson (1983, 1984a, 1984b, 1984c) has also investigated L2 writing development in bilingual children, although she has done so primarily through case studies. Both she and Edelsky (1982) found evidence of invented spellings in both the L1 and L2 writing samples of the children. This finding, again, echoes similar results found in the L1 literature.

Hudelson (1983, 1984a, 1984b) observes that children's writing is directly influenced by what they have read and/or what has been read to them. She also describes classrooms in which teachers advocate the teaching of writing through copying and emphasize correct form (Hudelson, 1983). Hudelson (1983) claims that in classrooms such as these, children learn that writing is copying and that risks should not be taken.

Findings on L2 writing development in children mirror findings on L1 writing development in children. Similarly, L2 research on composing with adults parallels much of the L1 research with adults. L2 studies on composing, like their L1 companions, provide valuable
insights into writers' different kinds of knowledge and the ways in which these kinds of knowledge influence writing.

L2 Composing Processes

Research on L2 composing processes began recently. Raimes (1985) notes that most of the studies of L2 composing processes are case studies involving a limited number of subjects. The predominant techniques used in process research are interviews, "think-aloud" protocol analyses, and self-reports. Raimes (1985) states that the "limited number of subjects makes it difficult to form conclusive generalizations. However, some patterns are emerging" (p. 231).

Similarities between L1 and L2 composing processes have been observed (Jones, 1982, 1983; Lay, 1982; Zamel, 1982, 1983). Zamel (1982, 1983) has described the stages of the L2 composing process (prewriting, drafting, revising, editing) and notes that they parallel the stages observed in the L1 composing process.

The literature on L2 composing processes also provides descriptions of the characteristics displayed by skilled and unskilled writers. These characteristics, again, are similar to those of skilled and unskilled L1 writers (Jones, 1982; Zamel, 1982, 1983). Skilled ESL writers tend to focus on generating language and are less distracted by errors in early stages (McKay, 1984; Raimes, 1985; Zamel, 1983). They view revision as re-seeing (Zamel, 1985) and editing as a process for clarifying ideas (Zamel, 1983). Skilled ESL writers also demonstrate their understanding of writing as creating meaning, which is evidenced by their willingness to be flexible and their desire to communicate to the reader (Zamel, 1982). Zamel (1983) has observed that skilled
writers develop strategies to avoid syntactic and/or semantic difficulties. Finally, skilled writers reread frequently (Zamel, 1982) and focus on what they intend to communicate (McKay, 1984).

In contrast to skilled ESL writers, unskilled ESL writers engage in very different strategies as they compose. Zamel (1983) explains that unskilled writers view writing in discrete parts (words, sentences, paragraphs), not as a creation of whole discourse. Furthermore, they appear to think that the writing process begins with a definite plan and that the writer should not deviate from that plan (Zamel, 1983). Both McKay (1984) and Zamel (1983) observe that less proficient writers are overly concerned with avoiding errors and that they attend to surface errors prematurely. Jones (1982) has noted that unskilled writers appear to be more constrained by the text and are unable to distance themselves from it, making it difficult to create "reader-based prose" (Flower, 1979). Finally, the revision strategies of less proficient writers are superficial; the writers rarely make changes that affect meaning (McKay, 1984; Zamel, 1983).

Raimes (1985) claims that researchers and practitioners must approach the above descriptions with caution—particularly those drawing close connections between unskilled L1 and L2 writers. She bases her warning on her observation that researchers often define "unskilled" differently. Raimes (1985) concludes that researchers must address this issue because "...generalizations from several studies can be made only if the same criteria are used" (p. 232). In fact, Raimes (1985) offers evidence to deny the assumption that L1 and L2 writers perform similarly. For example, the eight unskilled L2 writers in her
study were not preoccupied with error and with editing. They also reread frequently, which displays a strategy more associated with skilled writers. Raimes (1985) also notes that the strategies exhibited by her subjects were not consistent enough for her to be able to claim that they could be characterized as a group. She concludes, "these ESL writers revealed some similarities to [L1 unskilled writers] and to each other, but the patterns of behavior that emerged were not consistent enough for unskilled ESL writers to be described as a definable group..." (Raimes, 1985, p. 249).

Despite its shortcomings, research on L2 composing processes has led to some important pedagogical implications for ESL composition. There has been a call for more focus on process as opposed to product (McKay, 1984; Raimes, 1979, 1983; Taylor, 1981; Zamel, 1982). In addition, researchers have suggested that ESL students be given the opportunity to write journals (Spack & Sadow, 1983), to do written brainstorming and free writing (Taylor, 1981; Knepler, 1984), and to engage in peer response (Keyes, 1984; Stokes, 1984). Despite the research to support them, these pedagogical suggestions are not frequently followed in ESL composition classrooms.

The above statement is supported in part by evidence from L2 research on response to writing. Zamel (1985) notes that studies of teachers' responses in L2 settings similar to those conducted in L1 settings "are practically nonexistent" (p. 83). Hendrickson (1980) found that teachers' responses to errors are inconsistent and imprecise. Cumming (1983) observed that the technique used most frequently for responding to ESL writing is that of error
identification. His observation is echoed by Zamel (1985). Zamel (1985) conducted a study in which she examined 105 texts in order to analyze the responding behaviors of 15 ESL instructors. Her findings are consistent with those of the L1 response literature. Zamel (1985) concludes:

ESL writing teachers misread student texts, are inconsistent in their reactions, make arbitrary corrections, write contradictory comments, provide vague prescriptions, impose abstract rules and standards, respond to texts as fixed and final products, and rarely make content-specific comments or offer specific strategies for revising the text (p. 86).

It must be noted that Zamel (1985) does not indicate whether her subjects were selected from a variety of universities or from one. Her conclusions, therefore, must be accepted tentatively until they can be substantiated with further evidence.

Research on response does suggest, however, that research and theory have had a limited effect on practice in ESL classrooms. Over ten years ago, Zamel (1976) noted that too many instructional methods being used in ESL classrooms were not based on theory. Little has changed. According to a recent survey, ESL composition textbooks display a minimal influence of research and theory on practice (Raimes, 1986). Zamel (1987) concludes that "writing continues to be reduced to a set of discrete steps and prescriptive principles that students are exorted to follow in order to learn to write well" (p. 701).

The research on response to writing and on current pedagogical practices is important to writing assessment. The instruction students receive and the response given to their writing add to
the students' knowledge base about language and discourse components in writing. This knowledge will, in turn, have an affect on their writing performance under assessment conditions.

Summary

The above review of L1 and L2 composition research, in addition to research from other disciplines, has influenced the research that has dealt directly with topic variables. The vast majority of the research that deals with topic variables has been conducted in an L1 context.

Topic Variables

Central to the development of both testing and research instruments for directly assessing writing performance is the matter of creating appropriate topics that will elicit the responses necessary for an accurate assessment to be made. Devising effective topics is an exceedingly difficult and time-consuming task. White (1985) attributes this difficulty to the requirements that the topic (1) communicate the task clearly and succinctly, (2) elicit an immediate, appropriate response, and (3) elicit a wide range of creative responses. In answer to these requirements, researchers have devised criteria for creating effective prompts for the assessment of writing proficiency.

Suggested Criteria for Developing Topics

In the past two decades, lists of criteria have been suggested for the development of writing stimuli for testing purposes. These criteria can be summarized in terms of four general areas: wording, mode of discourse, rhetorical specification, and subject matter.
Many researchers have suggested that the wording of the prompt be clear and precise (Conlan, 1982; Hopkins & Stanley, 1981; Kirrie, 1979; Mehrens & Lehman, 1973; Ruth & Murphy, 1988; White, 1985). The stimulus must be constructed in such a way that the writer will produce a response that is representative of the kind of response the test designers are seeking.

The second general area for devising prompts includes criteria related to mode of discourse. Writing stimuli should provide the major organizing principle, e.g., compare/contrast, describe and analyze, etc. (Carlson et al., 1985; Conlan, 1982; Kirrie, 1979; Ruth & Murphy, 1988).

Criteria involving rhetorical specification constitute the third general area of suggestions. Some researchers contend that audience should be specified in the prompt (Farrell, 1969; Mellon, 1975; Ruth & Murphy, 1988; Sanders & Littlefield, 1975). Specification of purpose is also considered to be important (Farrell, 1969; Mellon, 1975; Irmscher, 1979; Lloyd-Jones, 1977; Sanders & Littlefield, 1975).

The final general area of suggested criteria for the development of writing stimuli is that of subject matter. Most researchers who have discussed criteria for task development maintain that the topic must stimulate interest in the writer (Carlson & Bridgeman, 1986; Conlan, 1982; Irmscher, 1979; Ruth & Murphy, 1988; White, 1985). Some of these researchers add that the subject matter of the prompt must also be interesting to the raters (Carlson & Bridgeman, 1986; Conlan, 1982; Ruth & Murphy, 1988; White, 1985). In addition, it has been suggested that topics allowing writers to take political, religious, or social
positions be avoided because the content of the responses may bias the raters (Carlson & Bridgeman, 1986; Conlan, 1982; White, 1985). Carlson and Bridgeman (1986), in relation to ESL writers, and Brick (1980, cited in Brown, 1986), in relation to Black writers, also emphasize the need to be aware of potential sociocultural bias in the content of the writing stimulus.

An issue related to the subject matter of writing stimuli is that of topic choice. Some test developers have offered a choice of topics on a test in the belief that students will be more free to write well if they are allowed to choose among several topics. Coffman (1971), Evans (1979), Freedman (1983), Ruth and Murphy (1988), and White (1985) argue that this belief is mistaken because some topics may be more difficult than others, and, thus, invalidate the scoring. White (1985) further claims that "it is much more fair to the students to offer choice within the format of a single topic" (p. 105). This argument is echoed by Beach and Bridwell (1984), who posit that, "if writers are not able to choose what they write about, they may not give their best performance" (p. 226).

Although the above proposed criteria for developing topics appear perfectly logical, they have not been derived from a strong research base. Ruth and Murphy (1988), in the most extensive review of research on task development to date, found "little research directly focused on the study of the properties of topics" (p. 50). Other scholars, such as Odell, Cooper, and Courts (1978), Odell (1979), and Hoetker (1982), also attest to the lack of knowledge about the development of topics
for writing assessment. The limited amount of research conducted thus far has provided important preliminary data that necessitate further investigation.

Research on Topic Effects

Identical to the four general areas of suggested criteria for the development of topics are the factors of prompts that have begun to be studied: wording, mode of discourse, rhetorical specification, and subject matter. Brossell (1986) and Ruth and Murphy (1988) have reviewed the research on these factors and have concluded that a clear understanding of the relationship of these factors to an individual's writing performance has yet to be achieved.

Wording. The scant amount of research on the wording of topics has produced inconclusive results. Studies conducted by Greenberg (1981) and Brossell and Hoetker Ash (1984) provided no evidence in support of the assumption that small changes in the wording of a topic affect writers or the raters' impressions of the writing. Ruth and Murphy (1988) argue, however, that this lack of evidence may be a result of the fact that the above studies were experimental in nature. By making use of both quantitative and qualitative data, a number of researchers cited by Ruth and Murphy (Keech & McNelly, 1982; Kinzer & Murphy, 1982; Leu et al., 1982) found differences among and between writers, raters, and test makers with respect to topic interpretation. Questionnaire and interview data revealed differences that post hoc comparisons of essays could not. Ruth and Murphy (1988) attributed these differences to factors involving the writers themselves (i.e., their prior knowledge, linguistic knowledge, and discourse knowledge), the text
itself (i.e., the combination of subject matter, instructions, and wording), and the testing context.

**Mode of Discourse.** Different modes of discourse required by writing stimuli produce varying degrees of syntactic complexity in the written responses. Several researchers (e.g., Cooper & Watson, 1980, as cited by Ruth & Murphy, 1988; Crowhurst, 1978; Crowhurst & Piché, 1979; Quellmalz et al., 1982) have concluded that narrative topics produce less syntactically complex responses than argumentative topics. In this same vein, other researchers (e.g., Neilsen & Piché, 1981) have found that certain topics appear to elicit responses that contain more sophisticated vocabulary and that this characteristic of the responses resulted in higher scores on the essays. According to Brosell (1986), several researchers (e.g., Nold & Freedman, 1977; Thomas & Donlan, 1982) have found that some topics also produce longer responses and that the length of the responses, in turn, appears to have a positive effect on the holistic ratings of the essays. Evidence in the literature also suggests that narrative forms are more easily acquired than abstract forms, such as those required in academic discourse (Freedman & Pringle, 1980; Keech, 1982; Matsuhashi, 1981). Matsuhashi (1982) posits that this difference may be due to the different cognitive demands placed on writers for different modes.

**Rhetorical Specification.** Studies on the provision of rhetorical specification in topics have led researchers to question the earlier assumption that topics containing full rhetorical specification are superior (Brosseil, 1983; Rowntree, 1977). Hoetker (1982) posits that topics containing full rhetorical specification may create more
opportunities for misinterpretation of the prompt to occur, which, in turn, would result in essays of poorer quality. Rowntree (1977) suggests a different explanation for what might happen to a writer as the level of specification is increased. He claims that full rhetorical specification restricts the topic so much that less variation in the responses results, making differences difficult to detect.

Several composition researchers (e.g., Calkins, 1983; Graves, 1979; Moffett, 1968) have identified audience awareness as an important factor in children’s writing development. Similarly, Flower and Hayes (1980) have demonstrated through the use of protocol analysis that a sense of audience is indicative of effective composing strategies. Audience specification has also been found to have an effect on syntactic complexity (Crowhurst & Piché, 1979; Rubin & Piché, 1979).

Research conducted under assessment conditions has produced inconclusive results with regard to audience and purpose specification (Brossell, 1986). Passe (1981) found assigned audience type to be a significant factor in holistic assessment. Witte et al. (1986) found that essays rated as high in quality were written on topics in which audience was specified and purpose unspecified.

**Subject Matter.** The subject matter of writing prompts has just recently emerged as a significant area in need of research. An issue related to the subject matter of the prompt is interest. A topic of interest to one writer may not be of interest to another. Faigley et al. (1981) claim that interest is of critical importance to students’ writing. Evans (1979), in a study with Canadian students, found that
"...the most serious problem with most of the writing was...a lack of personal commitment and interest" (p. 17). In contrast, Hoetker, Brosell, and Ash (1981) and Freedman (1983) reported that essays written on topics that were judged to be most popular among writers did not result in higher scores. Freedman (1983) found that although topics did affect scores, the students’ interpretation of the ease or the interest level of a topic was not the source of that effect. Freedman (1983) hypothesized: "topics that appear easy on first glance turn out, in fact, to be difficult. Likewise, topics that appear dull become interesting as writers explore their depths" (p. 323).

Related to the issue of interest is that of the writer’s familiarity with the subject matter of the writing stimulus. Several recent studies have explored the relationship between topic familiarity and writing.

Langer (1984) conducted the most comprehensive large-scale study on prior knowledge and writing to date. Four tenth grade American History classes, representing a total of 99 subjects, responded to two writing tasks at two points during the semester. The tasks were developed by the two history teachers who assisted in the study.

Langer first used a free-association task (Langer, 1980, 1981, 1982) to measure the subjects’ knowledge of the topics on which they were to write. On the basis of their responses to the free-association stimulus words, subjects were assigned three knowledge scores. The first score was for fluency, which represents the subjects' amount of available information for responding to the task. The second score, labeled organization, reflects the highest level of organization
achieved in the responses to the free association task. The third knowledge measure, referred to as combined knowledge, represents a combination of the fluency and organization measures.

Each writing sample was scored on five dimensions: overall quality, coherence, syntactic complexity, audience, and function. Data analysis was accomplished by means of Pearson product-moment correlations for examining the relationship among the knowledge and the first three writing measures. Audience and function measures were subjected to analyses of variance (ANOVAs).

Results of the correlational analyses indicated that the three knowledge measures were significantly related to the overall quality ratings of the samples. Only the combined knowledge measure was significantly related to the measure of coherence. Correlations among the knowledge measures and syntactic complexity measures, however, were not found to be significant.

Langer (1984) explained that a close analysis of the four teacher-developed topics revealed that the tasks required different writing strategies of the subjects. The topics follow:

1 Write a paper comparing city and frontier life with regard to individualism and democracy.

2 Write a one or two page essay on your version of a Utopian society, the kind you would like to live in.

3 It has been stated that in the 18th and 19th centuries the South was a deferential society. In one or two paragraphs, explain why this was true. In your answer, be sure to discuss the concepts of prejudice and acquiescence and how each related to this conclusion.
Some historians refer to the 1920s as a decade in American history when sexual freedom and the pursuit of happiness flourished. At the same time, it is noted that the 1920s were characterized by harsh moralistic and antiforeign sentiments.

Explain how social changes during the 1920s influenced the growth of new values that conflicted with traditional ones. (three fourths to one full page) (Langer, 1984, p. 31).

Two of the above assignments (#1 and #4) require a comparison/contrast pattern, whereas the other two (#2 and #3) require a thesis/support pattern. In addition, two of the prompts (#3 and #4) clearly provide more structure. Langer (1984) notes that "the type of argument required to respond appropriately to a prompt proved to have more influence on the complexity of the task than did the amount of structure provided by the prompt itself" (p. 36). Furthermore, for the comparison/contrast topics, the organization and combined knowledge measures were strongly related to essay quality. In contrast, for the thesis/support topics, the amount of information available (indicated in the fluency score) was related to essay quality, but the organization of that information was not. Langer (1984) posits:

These findings imply that different assignments, given for different purposes, tap different aspects of a writer's knowledge of a topic. A low score on a particular paper might not mean that a student "does not know the information" but that the information that was available was not organized in a useful way for that particular assignment (p. 40).

Audience type was examined by means of an ANOVA. Langer (1984) notes that some subjects addressed their essays to the teacher as part of an instructional dialogue. These subjects have significantly higher scores for the fluency measure than the subjects who addressed their essays to the teacher as examiner. The essays written by the
latter subjects, however, had higher scores for the combined knowledge measure and the measure of overall quality, although neither was significant. Langer (1984) posits that students who engage in a teacher–learner dialogue (a less formal approach than the one followed by subjects who addressed teacher as examiner) may do so because they lack sufficient knowledge of the topic to deal with it more formally.

The results of the ANOVA on function categories reveal similar patterns. The majority of the subjects wrote analytic essays, whereas a small number of subjects relied upon summary to compose their responses. Subjects who wrote analytic essays had significantly higher combined knowledge scores than subjects who summarized the information. Moreover, the subjects who had more organized knowledge available received significantly higher scores for overall quality. Langer (1984) suggests that when students have rather limited knowledge about a topic, "they may resort to writing summaries when analyses would have been more appropriate, because the summary format permits them to recount facts without having to interrelate them more fully" (p. 42).

Langer (1984) concludes that a strong, consistent relationship clearly exists between prior knowledge of topic and overall quality of writing. Correlational data, however, do not enable the researcher to draw definitive conclusions on the effects of prior knowledge. Langer's study, therefore, provides important preliminary data that need to be investigated more fully.

Like Langer's study, Cheskey's (1984) dissertation research focused on the effects of prior knowledge and audience on high school students' writing. Eight high school juniors participated at subjects in the
study. Forty subjects wrote on a topic about which they had low-prior knowledge (tobacco price supports), as demonstrated by their responses to a free-association task for measuring prior knowledge (Langer, 1980, 1981, 1982). Another 40 subjects wrote on a topic about which they were judged, on the basis of the prior knowledge measure, to have high-prior knowledge (the problems with teachers). Like in Langer's (1984) study, subjects were assigned three knowledge scores for fluency, organization, and combination. Half of the students in each prior knowledge group were instructed to write to their peers and the other half to their teachers.

The essays were scored on the basis of six measures: overall quality (as measured by holistic score), essay length, context-creating statements (i.e., statements provided by the writer to orient the reader to the essay), cohesion, syntactic complexity (as measured by T-unit length), and error analysis (as measured by the number of errors per 100 words). In addition, after the students wrote their essays, they rated their feelings toward their writing by responding to an involvement survey. On a five-point Likert scale, the subjects rated how involved they felt in the writing, how they liked their writing, whether they found the writing task easy or difficult, and if they wrote more for a grade or in order to say something to someone. Subjects were also asked to respond to an open-ended question about how and why they felt as they did about the writing task.

Analyses of variance were used to test for differences on the six writing measures. Cheskey (1984) explains that the combination score for prior knowledge was used in these analyses because it represents
the two highest levels of prior knowledge. Pearson product-moment correlations were calculated to determine the relationship among the six writing measures and the three knowledge measures for each subject.

Results showed that significant main effects for prior knowledge were found for holistic score, essay length, and context-creating statements. None of the other main effects or interaction effects, however, achieved significance. Significant correlations were also found for the three knowledge measures and holistic score, essay length, and context-creating statements. In sum, subjects with high-prior knowledge on the topic wrote qualitatively better essays, wrote quantitatively longer essays, and created more of a context for their readers than subjects with low-prior knowledge.

Although the ANOVAs did not result in significant effects for cohesion and error analysis, portions of the correlational analyses did achieve significance. A significant correlation was found between the organized knowledge measure and cohesion. Cheskey (1984) points out that this result suggests that the ability to organize prior knowledge helps a writer produce more cohesive writing. A significant negative correlation resulted between the combination knowledge measure and the error analysis score. Subjects with high combination scores made fewer errors.

No significance was achieved for either the syntactic complexity measure or the audience variable. Cheskey (1984) posits that other factors, such as mode of discourse, may have a greater effect on syntactic complexity than knowledge of topic. Crowhurst and Piché (1979) and Rubin and Piché (1979) have found that argumentative
discourse results in more syntactically complex writing than narrative discourse. These researchers also found that audience specification had a similar effect on syntactic complexity. Cheskey's (1984) results, however, do not lend support to earlier findings on audience. Cheskey (1984) hypothesizes that the lack of differences between the essays written for peers and those written for teacher was due to the content of the high-prior knowledge topic (the problems with teachers). It appears that this topic created a meaningful communicative situation for all subjects, regardless of the assigned audience. In other words, subjects given the high-prior knowledge topic had something to say to both audiences.

The results of the involvement survey indicate that subjects in the high-prior knowledge groups were more involved in their writing, liked their writing more, and found the writing task much easier than those in the low-knowledge groups. The item on the survey that involved "saying something to someone" did not result in a significant correlation with prior knowledge. Responses to the open-ended question confirmed the results of the survey "...the low-prior knowledge group made 27 negative comments and only six positive comments. In contrast, the high-prior knowledge group made only three negative comments and 21 positive comments" (Cheskey, 1984, p. 53).

Cheskey's (1984) study has provided the field with valuable insights into the effects of prior knowledge on writing. Of particular importance are the findings reported concerning the subjects' interest with the high-prior knowledge topic (the problems with teachers). It may be posited from the results of this study that interest in the
content of the writing stimulus is a more important factor than audience. Faigley, Daly, and Witte (1981), among others, certainly advocate the importance of interest. They argue that interest may be the most important influence on students' writing. The effects of interest on writing versus the effects of prior knowledge on writing will be a valuable area of inquiry in future research.

McCutcheon (1986) conducted a study on the effects of domain knowledge (i.e., prior knowledge of topic) and linguistic knowledge on the development of writing ability in children. The study involved 30 male children, ten each from fourth, sixth, and eighth grade. Five in each group were classified as high-knowledge subjects on the topic of football and five as low-knowledge subjects. This classification was made on the basis of the results of a 30-item completion test of the terminology and rules of football. The low- and high-knowledge subjects in each group were matched according to reading ability on the basis of their scores on the California Achievement Test.

The subjects generated a total of eight essays, four on the topic of football and four on their school or people they knew. The latter four represented the group of control topics. All subjects were assumed to be equally knowledgeable about the control topics. The subjects met individually with McCutcheon once or twice weekly and were asked to produce one text in each session. At these sessions, the subjects were given a topic sentence that contained one or more blanks and they were asked to use that topic as a framework for developing their texts. A sample topic was "I think that (some team chosen by the
writer) (will/won't) win the championship next year" (McCutcheon, 1986, p. 435). In order to motivate the children to put forth the necessary time and effort to create the texts, McCutcheon (1986) asked them to participate in a newspaper project. The subjects were told that their texts would be used for the sports and local news sections of the paper.

The essays were measured in terms of four dimensions: (1) coherence, as measured by the percentage of sentences judged to be coherent with each of a subject's eight essays, (2) length, as measured by the number of clauses in each essay, (3) elaboration, as measured by the mean number of clauses found in sentences that were judged to be elaborations of main points, and (4) content analysis, as measured by the percentage of clauses that discussed football at the level of specific play actions. The final measurement did not involve the control texts.

Results indicated that subjects produced more coherent texts on topics they knew well, regardless of grade level. Furthermore, when knowledge of topic was controlled, the texts increased in coherence as grade level increased: The older children wrote more coherent texts than the younger children. Subjects produced somewhat longer texts on topics they knew well, though length differences were not statistically significant.

High-knowledge subjects elaborated their texts more than low-knowledge subjects. In fact, control topics were more elaborated than football topics by the high-knowledge subjects. In contrast,
low-knowledge subjects elaborated their football and control texts equally. These results indicate that differences in elaboration were due to individual differences of the subjects, not to age or knowledge of topic. The content analysis showed that "high-knowledge subjects generated proportionally more clauses dealing with specific play actions than did low-knowledge subjects, 31.2% compared with 18.9%" (McCutcheon, 1986, p. 440).

McCutcheon (1986) concludes that the subjects' differences in linguistic knowledge best explain the observed grade level differences in the coherence analysis. She posits that, as children develop as writers, they gain more generalizable knowledge about texts and text production. This knowledge, according to McCutcheon, appears to be tapped by the coherence measure. McCutcheon further claims that although content knowledge does not necessarily compensate for immature writing strategies, linguistic knowledge can compensate for limited content knowledge. She bases this claim on a comparison of two texts. One text was written by a low-knowledge eighth-grade subject and the other text by a high-knowledge fourth-grade subject. The older child demonstrated the ability to elaborate and to produce cohesive ties between sentences, whereas the younger child relied on simply listing his ideas, even though his ideas were clearly more insightful in terms of content that were the older child's.

McCutcheon's (1986) argument that content knowledge does not compensate for a lack of linguistic knowledge is probably justified. It makes sense to believe that if a writer does not possess mature
discourse knowledge, that writer's content knowledge will not cause discourse knowledge to appear miraculously. McCutcheon's counter-argument that discourse knowledge can compensate for lack of content knowledge, however, is debatable. McCutcheon bases this rather strong argument on the results of the coherence measure alone. Coherence alone cannot determine the degree of an individual's discourse knowledge.

McCutcheon's (1986) study is important in that it adds to the limited data base on the effects of topic knowledge and discourse knowledge, two knowledge components discussed by many researchers, such as Bereiter and Scardamalia (1987), Flower (1988), and Purves and Purves (1986). McCutcheon's conclusions, however, are clearly in need of further investigation.

Although not initially developed to investigate the effects of prior knowledge of topic on writing, two other studies add to the growing literature base on prior knowledge of topic and its effects on writing. Hilgers (1982) conducted a study comparing the quality of college students' essays and their Scholastic Aptitude Test scores. He found small but consistent and significant correlations between the two (from .16 to .48, with a mean of .29, across nine measures). When familiarity of the content of the writing tasks was controlled for, however, the correlation decreased significantly (.04). Hilgers (1982) concluded that topic-specific prior knowledge may influence writing more than any particular training program or any research variable, such as general knowledge, ethnicity, previous education, or gender.
A year-long case study conducted at Carnegie Mellon University traced the writing progress of a first year graduate student in the rhetoric program (Berkenkotter et al., 1988). The researchers found that as the subjects’ familiarity with the issues of the rhetoric community increased, his writing improved. The subject produced increasingly more syntactically complex essays and more conventional expository patterns. Berkenkotter et al. (1988) note, however, that the subject "...still needs to develop ways of creating coherent and cohesive texts" (p. 36).

The above studies on prior knowledge were all conducted with subjects writing in their L1. The only L2 study on the effects of prior knowledge on writing to date was conducted by Winfield and Barnes-Felfeli (1982). These researchers investigated the effects of familiar and unfamiliar cultural contexts on the writing of ESL subjects. Ten Spanish-speaking ESL learners and ten ESL learners from a variety of language backgrounds were presented with two passages: one on the Spanish classic Don Quixote, and the other on the Japanese Noh Theater. After reading the passages, subjects were asked to produce immediate written recalls. The recalls were scored for overall length, grammaticality (on a five-point holistic scale), and mean T-unit length. The cultural familiarity of the Spanish-speaking group with the Don Quixote passage resulted in significant effects for overall length. Although grammaticality and mean T-unit length were also found to be higher in the Spanish-speaking recalls of the
contextually familiar passage, these measures were found to be statistically significant in terms of main effects only for groups, not for theme.

Summary

From the above review of research it is clear that writing research has made considerable progress. This progress, however, has had little effect on writing assessment, as evidenced by the discussion in Chapter One on current writing assessment procedures. It is imperative that test developers be more attentive to the findings of writing research. This is not to say that writing research findings are generalizable to writing assessment situations, but rather that findings from writing research should provide a framework for assessment studies. This is what the present study purports to do. This study is based on the research from cognitive psychology and L1 and L2 reading, and on the emerging literature base from L1 and L2 composition research. Its intent is to apply this research base to an L2 writing assessment situation.

In the present study, the L2 subjects’ writing performance will be assessed on the basis of the essay responses they compose. Data in the form of essays that are written by subjects in an assessment situation can be scored in a variety of ways. The scoring procedures chosen for the present study are based on the following review of the research on conventional methods for scoring compositions.
Conventional Methods for Scoring Compositions

Holistic Score

The holistic scoring procedure, defined as a quick, impressionistic qualitative procedure for sorting or ranking samples of writing, has been in widespread use in both L1 and L2 research. Central to any evaluation procedure are the issues of reliability and validity; holistic scoring has received its share of criticism and support on both counts.

The reliability of holistic scoring has perhaps received the majority of researchers' attention. On the one hand, a number of researchers attest to the reliability of this procedure (e.g., Braddock et al., 1963; Pollman & Anderson, 1967; Kaczmarek, 1980; Moslemi, 1975). Several researchers, on the other hand, have provided evidence against the reliability of holistic scoring (Coffman, 1971; Cronbach, 1971; Diederich, 1974; Diederich et al., 1961). Inconclusive results have been reported by Mullen (1980). Perkins (1983) speculates that lack of reliability of holistic scores may be a result of "bias, fatigue, internal lack of consistency, previous knowledge of the student, and/or shifting standards from one paper to the next" (p. 653).

In an attempt to control for these factors that may affect reliability, several researchers have suggested certain criteria to be followed when evaluating essays holistically (Braddock et al., 1963; Cooper, 1977; Godshalk et al., 1966; Henning, 1984; Jacobs et al., 1981). White (1985) succinctly summarizes these criteria as follows:
1 **Controlled Essay Reading.** All raters are brought together to read and score the essays at the same time and place with the same working hours and breaks.

2 **Scoring Criteria Guide.** A direct statement of descriptors for papers at different points on the scoring scale is developed by those leading the controlled essay reading before the readers gather.

3 **Sample papers.** These papers, which are chosen by the leaders of the reading, provide examples of papers at the points of the scale and are given unmarked to readers for scoring during the training session.

4 **Checks on the Reading in Progress.** Repeated checks are orchestrated by the leaders in order to reinforce the group standards and help keep individuals from drifting from original standards.

5 **Multiple Independent Scoring.** Two different readers score each essay independently. A one-point difference between scores is acceptable; a difference of two or more points necessitates a third reading, which resolves the discrepancy.

6 **Evaluation and Record Keeping.** A continuing program retains records of scoring done by readers in order to keep the most reliable scorers for subsequent readings (pp. 24-27).

Jacobs et al. (1981) also suggest that raters come from similar backgrounds. One final criterion offered is that raters read quickly (Cooper, 1977; Godshalk et al., 1966).
Similar to the discussion around the issue of reliability have been the arguments in favor of and against the validity of the holistic scoring procedure. Those researchers who support the validity of holistic scoring include Flahive and Snow (1980) and Kaczmarek (1980), whose research provides evidence for the concurrent validity of holistic scoring. In addition, Perkins (1983) claims that "holistic scoring has the highest construct validity when overall attained writing proficiency is the construct to be assessed" (p. 652). Similarly, Cooper (1977) asserts that "holistic evaluation of writing remains the most valid and direct means of rank-ordering students by writing ability" (p. 3). These latter two researchers, however, do not provide empirical support for their claims.

In contrast to the above assertions, Charney (1984) offers a counter-argument on the grounds that the "requirements for achieving reliable results from holistic ratings...bring the validity of the ratings into doubt" (p. 65). Charney (1984) notes that some researchers may argue that holistic scoring is invalid from the start because it is a product-based method of assessment and does not take the writing process into account. This view is extreme, however, and, as Charney (1984) points out, "product-based evaluations are the only feasible methods available for testing and for research projects involving large numbers of writers" (p. 68). Charney primarily bases her argument against the validity of holistic scoring on two points: (1) raters are often influenced by superficial features of the writing, such as poor handwriting, and (2) a generally accepted set of valid criteria by which raters score the essays has yet to be devised.
In short, she argues that until researchers come to a consensus on the qualities of good writing, it is unwise to assume that any set of criteria used to judge the essays can be considered valid. White (1985) mentions these concerns and adds that the holistic scoring procedure is also limited because of its inability to provide (1) diagnostic information, (2) absolute values that can be used each time holistic scoring is used, and (3) unquestionable reliability. White (1985) concludes, however, that despite its limitations, holistic scoring is "the most successful method of scoring writing in quantity that is now available" (p. 30).

**T-unit Measures**

The T-unit was developed by Hunt (1965) as an index of syntactic maturity in the writing of children and adolescents. T-units are defined as "minimal terminable units...since they [are] minimal as to length, and each...grammatically capable of being terminated with a capital letter and a period" (Hunt, 1965, p. 21). Since its development, the T-unit has been used as a research tool to measure the overall syntactic complexity of speech and writing samples in L1 research (e.g., Loban, 1976; Mellon, 1969; O'Donnell et al., 1967; O'Hare, 1973) and, more recently, in L2 research (Arthur, 1979; Cooper, 1976; Flahive & Snow, 1980; Gaies, 1976; Kameen, 1979; Larsen-Freeman, 1978; Larsen-Freeman & Strom, 1977; Monroe, 1975, Perkins, 1980; Perkins & Homburg, 1980; Vann, 1979; among others). The T-unit has not, however, been without criticism.

The criticism on T-units stems primarily from the emphasis the measure puts on the surface structure of an utterance (Moffett, 1968;
Ney, 1966). Researchers have pointed out that there is no necessary connection between syntactic complexity and the overall communicative effectiveness of a writer’s or speaker’s meaning (Gaies, 1980; Hirsch, 1977). Researchers are, therefore, cautioned against using an index based on syntactic complexity alone as a measure of overall language proficiency without consideration of vocabulary, stylistic effectiveness, appropriateness, and overall communication of meaning. It is important to keep in mind, however, that more proficient language users do show the ability to produce more syntactically complex utterances and that growth in this ability provides one effective index of language development that can be used in conjunction with others.

Specific to the use of the T-unit as a tool in L2 research is the issue of errors. Second language learners will produce errors in the target language and an index of L2 development must reflect this occurrence of errors. This issue has led L2 researchers to develop a modification of the basic index, mean T-unit length. "The length of error-free T-units is now considered to be a more valid measure of growth in a second language" (Gaies, 1980, p. 55, his emphasis). With this new index, another problem develops; what constitutes error?

Different researchers have defined error in various ways. Larsen-Freeman and Strom (1977) in their study of ESL compositions, for example, demanded that in order for a T-unit to be judged error-free, it had to be perfect in all respects, including spelling and punctuation. A less stringent definition was offered by Scott and Tucker (1974), who considered a T-unit to be error-free if it did not contain morphological or syntactic errors. Vann (1979), taking an
intermediate position, required that a T-unit be free of morphosyntactic and lexical errors, and that it make contextual sense in order for it to be deemed as error-free. Until researchers agree upon the definition of an error-free T-unit and establish a hierarchy of errors to reflect the fact that different errors have different effects (Gaies, 1980), what appears to be important to research is the consistency with which the T-units are judged to be error-free.

Some conflicting results have been reported in terms of the power of the error-free T-unit measures to discriminate among learners in adjacent levels of proficiency (e.g., Cooper, 1976; Larsen-Freeman & Strom, 1977; Monroe, 1975). According to other researchers, however, the error-free T-unit and its mean length have been found to provide valid indicators of different levels of language proficiency (e.g., Cooper, 1976; Gaies, 1976; Kameen, 1979; Larsen-Freeman & Strom, 1977; Monroe, 1975). Similarly, Perkins (1980) found that both these measures were highly significant in discriminating among different holistic evaluations of compositions from an advanced level.

**Overall Length**

Length, as measured by the number of words per essay, has been reported by some researchers to increase proportionally with levels of proficiency. Larsen-Freeman and Strom (1977), for example, reported a significant correlation between length and proficiency, and cite Neuman (1977) as having found comparable results. In a similar vein, Breland and Jones (1982) concluded that length influences reader judgments in the holistic evaluations of essays and hypothesized that this relationship occurred "probably because good discourse requires length
for proper development" (p. 28). Perkins (1980) investigated length to
determine whether it discriminated between different holistic
evaluations from one level of proficiency. He reported a slight
increase in length as scores on the essays increased, but the increase
was not statistically significant. Parallel findings were reported in
studies on the effects of topic familiarity.

McCutchen (1986) studied the effects of topic familiarity on
children's writing and found that children produced somewhat longer
texts on topics they knew well. In a study on the effects of topic
familiarity on the writing of high school students, Chesky and Hiebert
(1987) echoed McCutchen's results and furthermore found them to be
statistically significant. They concluded that "apparently more
coherent knowledge about a subject allows the writer to express more
about the topic" (p. 309). Similarly, in a study involving topic
familiarity and its effects on the writing of L2 learners, Winfield and
Barnes-Felfeli (1982) found that subjects familiar with the topic wrote
a much greater number of words on that topic than on the topic with
which they were not familiar.

Conclusion

As evidenced by the previous review of literature and the
description of assessment practices in Chapter One, a gap exists
between current writing assessment practices and the theory and
empirical evidence that have emerged from L1 and L2 research on
writing. The present study attempts to bridge this gap. The study
investigates the extent to which L2 students' writing is affected
by their knowledge of the subject matter of the topic used to elicit their writing performance. Their writing performance was judged on the basis of overall quality, syntactic complexity, and fluency. These judgments were made by means of the conventional methods used for scoring compositions: holistic ratings, counts of T-unit indices, and measures of length.
CHAPTER III

DESIGN AND PROCEDURES

Population and Sample Selection

The population from which the sample was drawn consisted of international graduate students enrolled in composition courses offered by the English as a Second Language (ESL) Programs at The Ohio State University (OSU), a state-supported, land-grant institution. All international graduate students who are accepted to this university (generally with a TOEFL score of 500 or better) are required to take a written placement examination with the ESL Programs. The examinations are evaluated by trained ESL staff members. The results of the examinations determine either that the student demonstrates a writing ability sophisticated enough to qualify, i.e., to place out of the courses, or that the student is in need of instruction in composition. In the latter case, the student must register for a class in the series of ESL courses offered. Students who necessitate instruction are placed, on the basis of the examination results, into a beginning, intermediate, or advanced level of ESL composition. Students representing these three levels were selected for this study.

For purposes of this study, the decision was made to select only those students who were new to the United States and to OSU. This stipulation was made by the researcher in order to obtain a sample
that was as homogeneous as possible in terms of previous experience in an English-speaking country. This information was obtained by means of a personal data questionnaire that is administered to all students in the ESL courses. Any students who had had previous university experience in an English-speaking country were not included in the study.

All of the subjects who participated in the study were volunteers. They were previously informed that their writing was to be used for research purposes and that they had the option to inform the researcher not to include their writing in the sample. A total of 105 subjects participated: 43 each at the beginning and intermediate levels and 19 at the advanced level. There were fewer subjects in the advanced group because most students at this level had previously taken one or more of the ESL courses offered and, therefore, did not meet the researcher's criterion that the subjects be new to the United States and OSU. The subjects chosen for the sample represented the following native language groups: Chinese, Korean, Japanese, Burmese, Indonesian, Spanish, French, Portuguese, German, Serbo-Croatian, Thai, Greek, and Arabic.

Research Design and Variables

For the purposes of statistical analysis, a one between-subjects one within-subjects mixed design was used for the experiment (Figure 1). This design is a mixture of the one-factor completely randomized and the one-factor repeated measurements designs. In such a design
Variable B: PROMPT TYPE

<table>
<thead>
<tr>
<th>General Prompt</th>
<th>*</th>
<th>Field-Specific Prompt</th>
</tr>
</thead>
</table>

Variable A: *

COURSE LEVEL *

1 Beginning (n=43) *

2 Intermediate (n=43) *

3 Advanced (n=19) *

Figure 1: One between-subjects (Course Level) one within-subjects (Prompt Type) mixed design.
each subject receives all levels of the within-subjects variable. The first independent variable, a between-subjects variable, was Course Level and consisted of three levels:

1. Beginning (English 106)
2. Intermediate (English 107-G)
3. Advanced (English 108.02)

The second independent variable employed in this study was Prompt Type, a within-subjects variable containing two levels:

1. General Prompt
2. Field-Specific Prompt

Six dependent variables were evaluated in the study. They were:

1. Holistic Score
2. Number of T-units
3. Mean Length of T-units
4. Number of Error-Free T-units
5. Mean Length of Error-Free T-units
6. Overall Length

On the basis of the research findings on the holistic scoring procedure discussed earlier, the decision was made to use this procedure in the present study. Because the primary focus of this study was on the differences between essays written by the same subject, holistic scoring was an effective measure of the overall quality of the compositions. Overall impression marking was the actual procedure followed during the holistic scoring session because the researcher's intent was to duplicate the normal procedures followed by instructors in the ESL Programs when scoring essays holistically.
T-units, error-free T-units, and their respective mean lengths were chosen to serve as measurements for the syntactic complexity of the essays. This choice was based on the research reviewed earlier. Although T-unit indices have been criticized, they were considered to be effective indices of syntactic complexity for the present study because the emphasis of the study was on the differences between repeated writing measures by the same subject.

The previously mentioned studies involving topic familiarity were most influential in the researcher's decision to include overall length as a dependent variable. In the present study, length provided an indication of the subjects' fluency with regard to the topic of the different prompts. In other words, length provided the means by which the researcher determined whether subjects demonstrated that they had more to say about the topic with which they were familiar.

Materials

The stimuli used in the present study consisted of two prompts to which subjects were asked to respond in writing. The first prompt was a question of general nature (Appendix A). This prompt was chosen from the repertoire of topics used by the ESL Programs for the diagnostic essay administered at the beginning of each quarter. The goal of the researcher in choosing the prompt was to find a topic that would require the subjects to compare and contrast the topic and take a position. This decision was made on the basis of a survey of university professors taken by Bridgeman and Carlson (1984) of the Educational Testing Service (ETS) in an attempt to determine the types
of writing tasks faced by beginning undergraduate and graduate students. According to the survey, this prompt type (comparison/contrast plus take a position) rated among the topics judged to be most acceptable by professors of graduate students.

The second prompt, created by the researcher, was a question pertaining to the subjects' field of study (Appendix B). One of the researcher's goals in developing the topic was to devise a topic similar in type to the first prompt (i.e., comparison/contrast plus take a position). It was also of interest to the researcher to develop an open-ended prompt, i.e., one that would allow the subjects to write on a topic of their choice that pertained to their field of study. This feature of the prompt enabled the researcher to assume that subjects were writing on a topic with which they were familiar.

A personal data questionnaire was used to determine whether the subjects were new to the United States and OSU (Appendix C). This questionnaire was designed by the ESL Programs and is administered each quarter to all students enrolled in the ESL courses.

Experimental Procedures

The First Phase

The experimental procedures for the study were carried out in two separate phases in which repeated writing measures were obtained from each subject. The first phase corresponded with the administration of the ESL diagnostic examination for Fall Quarter, 1987. At the beginning of each quarter, students in the composition courses are provided with the opportunity to write an in-class essay, called the
diagnostic examination, which is similar to the placement examination. The results of the diagnostic examination determine whether new students have been placed into the correct level and whether students previously enrolled in one of the courses have progressed substantially to be able either to pass into a higher level in the series or, in the case of advanced students, to qualify.

Prior to the start of the quarter, the researcher met with all 106, 107-G, and 108.02 instructors in order to explain the procedures for the administration of the diagnostic examination. The instructions (Appendix D) were explained in their entirety at the meeting and all instructors agreed that they had fully understood the procedures they were to follow. The instructors announced to their classes on the day before the examination that the students would be asked to write an in-class essay, which could potentially serve to move some students to a higher level in the series or to qualify students. Neither the students nor the instructors were informed of the topic of the essay prior to the examination.

On the day of the examination, Friday, September 25, the instructors were given large envelopes containing a copy of the instructions for administering the examination (Appendix D), a page for comments regarding unanticipated events (Appendix E), extra paper, and the examination packets. Each packet contained two sheets of lined paper, which were labeled with a number, and a folded sheet containing the general prompt, which was sealed and labeled with the same number. Subjects were asked to fill in the information requested on the outside of the folded sheet (Appendix F), and at the sound of the bell for
class to begin, the subjects were instructed to open the folded sheet. The instructors read the prompt aloud and told the subjects to begin the examination. At the end of 48 minutes, the essays were collected by the instructors.

Upon returning to the ESL offices, the instructors then read the essays and recommended to the directors those students who they believed should be moved to higher levels. The directors then, as per normal diagnostic exam procedures, decided which students were to move to a higher level or to qualify. All essays were then given to the researcher along with the personal data sheets of the students who had taken the examination. The data sheets provided the researcher with the information needed to determine which students met the criteria for the study. Once this task was accomplished, the researcher compiled a list of subjects, their fields of study, their ESL course levels, and their corresponding examination numbers. All essays were kept by the researcher. Students who had originally placed into one level but were moved into a higher level on the basis of the diagnostic essay, were considered to belong to the higher level into which they were moved.

The Second Phase

The second phase of the experiment took place three weeks later on Friday, October 16. Prior to the second phase, all 106, 107-G, and 108.02 students received a memorandum (Appendix G) informing them that they would be asked to write a second essay, similar to the one they wrote for the diagnostic examination, which was for research purposes only. They were told that the results of this second writing exercise, unlike the results of the diagnostic examination, would in no way
affect their status in the ESL program. The procedures followed for the second phase were identical to those of the first except that instead of a general prompt, the examination packet contained the field-specific prompt. (See instructions in Appendix H.) After administering the writing exercise, the instructors submitted the essays to the researcher; the researcher then found the essays written by those subjects who were selected after the first phase on the basis of the criteria of the study. Several of these original subjects either had dropped out of the courses or had not appeared for the second phase of the study. These subjects, therefore, were excluded, leaving a total of 138 subjects, 43 at the beginning level, 76 at the intermediate, and 19 at the advanced. After consulting with the Statistics Department at OSU, the researcher decreased the large number of subjects at the intermediate level. Using a table of random numbers (Gay, 1981), the researcher selected 43 of the 76 subjects at that level, leaving a grand total of 105 subjects.

Data Collection

Holistic Scores

Preliminary Matters. Eight experienced raters, three from ESL compositions courses, three from the English Department, and two from the American Language Program (ALP), a branch of the ESL Programs that offers an intensive language program, were asked to participate in the holistic scoring procedure. These raters were informed that the results were to be used in dissertation research, but they were not informed of the nature of the research. For their participation the
raters were each paid $35.00, which was made available from research funds provided by the ESL Programs.

Prior to the holistic scoring session, 24 essays (Appendix I) were selected as samples for the calibration session. The samples were chosen from the repertoire of essays authored by students who were not chosen as subjects for the study and those essays used for calibration purposes in the pilot study. Four essays, two in response to the general prompt and two in response to the field-specific prompt, were selected to represent each of the six levels on the scale. The selection was made by means of a collaborative effort on the part of the researcher and the director of the ESL Programs.

At the holistic scoring session, the eight raters met with the researcher in a large ESL office. They each sat at separate desks in the same room and were provided with a copy of the two prompts, a scoring sheet (Appendix J), and a pencil. They were informed that they would be reading essays in response to the two different prompts that were written by international graduate students and that they would be rating the essays on a scale from one to six, six being high on the scale, on the basis of their overall impression of the essay. This overall impression was to be a result of their taking into account such factors as organization, syntax, content, vocabulary, coherence, etc. These factors were discussed at length in the calibration session to follow. Raters were instructed not to make any marks whatsoever on the essays themselves and were told to mark their scores next to the numbers of the essays on the scoring sheet provided. Raters were
finally informed that each essay would be read twice and that essays receiving a difference in scores of two or more points would require a third reading.

In addition, the raters were asked to supply informative notes regarding factors of the writing that were influencing the raters' individual impressions. In the event that raters judged an essay to be off topic, they were instructed to score the essay based on their overall impression of the writing and to mark "OT" for "off topic" next to the score. In addition, raters were asked to make note of essays that were difficult to read due to poor handwriting by marking "HW" for "handwriting" next to the score. The last piece of additional information requested by the researcher was each rater's indication of his or her familiarity with the content of the written response to the field-specific prompt. Raters were asked to circle the "Y" on the scoring sheet next to the number of the essay they were reading if they judged themselves to be familiar with the content and to circle "N" if they did not. This additional information was requested by the researcher because it represented extraneous factors that could be responsible for the potential occurrence of discrepant scores among pairs of raters.

The Calibration Session. After the preliminary matters were discussed, the calibration session began. This session lasted approximately one and a half hours. The researcher began the calibration session by distributing xeroxed copies of two essays, which represented a sample of responses to both prompts at the sixth level of the scale. (See numbers 258 and 411 in Appendix I.) The raters
were told that each level was to reflect a range of scores; in other words, within each level it was possible to have low, average, and high essays. The next two samples distributed represented essays at the fifth level of the scale. (See numbers 403 and 80 in Appendix I.) After reading these samples, the raters were asked to compare the level six samples with the level five samples and to determine whether they saw a clear distinction between the two levels. Calibration progressed in this manner and discussion of essay characteristics succeeded the comparisons of essays at adjacent levels.

Upon finishing the discussion of samples representing all levels of the scale, the raters were given three unmarked essays and asked to read and rate them one at a time. (See numbers 729, 179, and 08 in Appendix I.) The researcher then compared the raters’ scores and found that close agreement (generally within one point on the scale) on the scores of these sample papers was achieved. In this way, the researcher determined that all raters were calibrated according to the original standards established by the sample essays.

Plan for Systematic Multiple Independent Scoring. In order to achieve systematic multiple independent scoring and to facilitate the recording process, the researcher had previously randomly assigned a letter (A-G) to each of the raters. In addition, the researcher had randomly distributed the 210 essays among 25 folders so that each folder contained approximately the same number of responses to each prompt. The researcher then used the letters assigned to the raters as a guide for creating 24 different pairs of raters and for equally distributing the first 24 folders among the pairs for first and second
readings. (See list of pairs and corresponding folders in Appendix K.) For example, rater A was the first reader for essays in folders numbered one, nine, and 17. These folders were later given to raters B, D, and G, respectively, for the second reading. This organization ensured that each folder, containing eight or nine essays, was read by a different pair of raters. The researcher was responsible for providing the readers with the correct folders. The raters were simply asked to place finished folders on the chairs next to their desks. The twenty-fifth folder, which contained eight randomly selected essays, was read by all raters for the purpose of an additional calibration check.

The Scoring Session. At eleven o’clock in the morning raters began the first reading of the essays in the folders that had been placed on their desks. Approximately one hour into the scoring session, raters were given two unmarked essays to rate in order for the researcher to check calibration. This practice was repeated a total of three times throughout the rating session in order to periodically ensure that raters remained calibrated with the original standards. One final calibration check, about which the raters were not aware, occurred during the session. Prior to the scoring session, the researcher and the ESL director randomly selected 20 of the 210 essays and rated them according to the calibration samples. The numbers of these essays and the ratings given to them were recorded by the researcher and were later compared with the scores awarded to the same essays during the actual rating session (Appendix L).
As the second readings were completed, the researcher recorded the scores given each essay. Those essays with scores showing a difference of two or more points were read by a third rater. The actual holistic score awarded each essay represented the sum of the two scores given by the raters. In the case of essays requiring a third reading, the total score reflected a sum of the closest two out of three scores, or, in the cases where the third rater’s score fell exactly in between the first two scores, the total score was represented by the third rater’s score doubled. For example, essay #592 received a score of two from rater F, a five from rater C, and a four from rater G, making its total score a nine, or the sum of the second and third raters’ scores. Similarly, essay #301 received a one from rater F, a three from rater A, and a two from rater D, making its total score a four, or the third rater’s score doubled. For those essays in folder 25, which were rated eight times, the researcher determined the mode and doubled it in order to arrive at the total holistic score.

Fourteen essays were not read by a third rater because the rating session lasted longer than anticipated and several raters had time conflicts. These 14 essays, however, received scores having a discrepancy of only two points. The researcher, therefore, summed the two scores in order to create their total holistic scores. Before the total holistic scores were subjected to statistical analyses, the researcher calculated interrater reliability for the pairs of scores awarded to the essays. The Pearson product-moment coefficient achieved a level of 0.8584.
T-unit Measures and Overall Length

Two staff members of the ESL Programs, who were experienced with the scoring of essays on the basis of their T-units, were asked to participate in this portion of the study. Each was paid $125.00, which was made available from research funds provided by the ESL Programs. The raters met with the researcher for a 45-minute session during which the researcher explained that the raters were to separate the essays into their T-units and determine which were error-free. Error-free T-units were defined as those T-units that did not contain syntactic or morphological errors, or errors pertaining to article usage. Spelling and punctuation errors were permitted as well as some lexical errors. These latter errors were left to the judgment of the raters. The researcher also asked that the raters verify that T-units made sense within their particular context in order to consider them to be error-free.

After this explanation, the researcher and the raters independently separated four sample essays into their T-units and determined which were error-free in accordance with the criteria established by the researcher (Appendix M). Discussion succeeded the rating of each essay and consensus was reached among the raters and the researcher. Each rater was given xeroxed copies of the 210 essays and asked to mark a slash (/) between T-units and to highlight those T-units that they considered to be error-free. The raters were allowed six weeks in which to score the essays independently.

After the essays were marked, they were returned to the researcher, who compared each pair of essays in order to verify the number of
T-units and error-free T-units. Interrater reliability coefficients were calculated using Pearson product-moment correlations. The degree of interrater reliability achieved for T-units was 0.9961 and for error-free T-units was 0.9159. In addition to totaling the number of T-units and error-free T-units per essay, the researcher calculated the mean number of words per T-unit and per error-free T-unit in each essay. Finally, the researcher counted the total number of words in each essay. This measure served as the sixth dependent variable, overall length.

Data Analysis

The six dependent variables were subjected to a multivariate analysis of variance (MANOVA) using the Statistical Analysis System (SAS) located on the mainframe computer at the Instruction and Research Computer Center at OSU. The data were also subjected to discriminant analyses (DAs) on the effects detected in the MANOVA and to separate two-way univariate analyses and post hoc procedures using Scheffé's test of significant differences. The following null hypotheses were tested:

As demonstrated by dependent measures of holistic score, number of T-units, number of error-free T-units, mean length of T-units, mean length of error-free T-units, and overall length:

Ho1: There was no significant difference between responses to the general prompt and field-specific prompt.

Ho2: There was no significant difference among the three course levels.
Ho3: There was no significant interaction between the prompt variable and the course level variable.

Pilot Study

Pilot testing of the experimental procedures took place during Summer Quarter, 1987. The subjects in this study were graduate students enrolled in the advanced ESL composition course (108.02). Seven raters participated in the holistic scoring session, which was based on a scale of one to four because the responses being rated represented only one course level. The researcher alone determined the number of T-units and error-free T-units, the mean length of T-units and error-free T-units, and the overall length of the essays. After careful scrutiny of all phases of the experiment, the researcher concluded that the prompts elicited appropriate responses and that the experimental procedures were generally acceptable. Minor modifications were made to the instructions given to the ESL instructors, to the memorandum given to the subjects, and to the procedures during which the data were collected.
CHAPTER IV
RESULTS AND DISCUSSION

Introduction
A great deal of research evidence from the fields of cognitive psychology and L1 and L2 reading supports the notion that prior knowledge affects the way individuals perceive and comprehend events and reading passages. This notion has recently been investigated in writing research. Prior knowledge of the topic of a writing stimulus has been found to influence writing performance in critical ways. The purpose of the present study was to apply the research base from the above disciplines to investigate L2 writing in an assessment setting in order to explore the effects of topic familiarity on L2 writing performance. In addition, this study sought to provide baseline data on ESL writers engaged in graduate study.

To assess the effect of topic familiarity over different groups of ESL writers, one independent variable, course level, was divided into three levels: beginning, intermediate, and advanced. The second independent variable, prompt type, had two levels: the general prompt, a question of general nature that was used to elicit the writing performance, and the field-specific prompt, a question pertaining to the subjects' field of study that was used to elicit the writing performance. All subjects at each of the three course levels responded
in writing to both prompts. The written essays were scored on the basis of six criteria, which served as the dependent variables: holistic score, number of T-units, mean length of T-units, number of error-free T-units, mean length of error-free T-units, and overall length. The holistic scores were determined by a team of eight experienced raters. The T-units and error-free T-units were determined by two expert raters. The remaining dependent variables were calculated by the researcher.

The six dependent variables were subjected to several analyses. When a number of variables are involved in a study, they first must be observed simultaneously so that the contribution of each variable to the overall results can be assessed. A multivariate analysis of variance (MANOVA) provided for this simultaneous analysis in the present study. The MANOVA results were then interpreted by means of a Discriminant Analysis (DA). The DA provides additional information about the nature of the differences found in the MANOVA. Significant MANOVA effects can be interpreted further by means of separate univariate analyses of variance (ANOVAs) on the dependent variables. Hummel and Sligo (1971) indicate that significant MANOVA results permit the researcher to continue with an interpretation of separate ANOVAs with the confidence that the overall alpha level will be under control. In other words, the MANOVA prevents possible misinterpretations of significant results of univariate ANOVAs on several variables by first examining all the variables simultaneously. According to Stevens (1986), subsequent investigation of the results of univariate analyses has greater power for detecting differences than
other post-hoc procedures suggested for the interpretation of the MANOVA. Consequently, the six dependent variables in the present study were also subjected to univariate ANOVAs. In addition, Scheffé's test of significant differences was used to make post-hoc comparisons when warranted.

The results of the MANOVA will be reported first. This report will be followed by a report of the results of the DAs on group and topic effects. Subsequent reports on the ANOVAs and post-hoc tests for each of the six dependent variables will provide a thorough analysis of the data, which will be followed by the discussion.

Results of the Multivariate Analysis of Variance

The results of the two-way MANOVA (Table 1) revealed highly significant main effects for both the course level variable, \( F(12, 194) = 8.76, p < 0.0001 \), and the topic variable, \( F(6, 97) = 10.80, p < 0.0001 \), when all six dependent variables were analyzed simultaneously. No interaction between the course level variable and the topic variable, however, was detected by the MANOVA, \( F(12, 194) = 1.69, p > 0.0718 \). The MANOVA only reveals a difference among the groups and between the prompt types. The nature of these differences must be determined on the basis of an examination of the DAs and univariate ANOVAs.

Results of the Discriminant Analyses

In order to explore the nature of the significance of both the course level and topic main effects detected in the MANOVA, DAs were
<table>
<thead>
<tr>
<th>Source</th>
<th>Wilk's Lambda</th>
<th>df N</th>
<th>Rao's F Test</th>
<th>Approximation</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between Ss</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Level (A)</td>
<td>0.4208</td>
<td>12.00</td>
<td>194.00</td>
<td>8.76</td>
<td>0.0001</td>
</tr>
<tr>
<td>S(A)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Ss</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic (B)</td>
<td>0.5995</td>
<td>6.00</td>
<td>97.00</td>
<td>10.80</td>
<td>0.0001</td>
</tr>
<tr>
<td>AB</td>
<td>0.8199</td>
<td>12.00</td>
<td>194.00</td>
<td>1.69</td>
<td>0.0718</td>
</tr>
<tr>
<td>SB/A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
performed. The DA determines the maximum discrimination among the
groups in terms of individual subjects' scores on the six dependent
variables when the dependent variables are viewed simultaneously. The
DA calculates a number of discriminant functions according to the number
of levels of the independent variable. The number of discriminant
functions equals the number of levels of the independent variable minus
one. The significance of a particular function is determined on the
basis of its degree of total discriminatory power, as indicated by the
Eigen Value. A significant function serves the purpose of explaining
the proportion of contribution each dependent variable makes to the
overall discriminatory power of the function. As seen in Table 2, of
the two discriminant functions calculated for course level effects,
only function one was found to be significant. The nature of this
significance is determined by examining the relative magnitude of the
structure coefficients for the function. These magnitudes reveal the
relative importance of the dependent variables. Structure coefficients
having values greater that 0.30 are traditionally regarded as
meaningful (Pedhazur, 1982).

Table 2 shows that function one accounted for 96.3% of the total
discriminatory power. The major variable contributing to this
discrimination of groups according to course level was the holistic
score (0.7827). Other variables contributed in similar degrees to the
discrimination of groups, though to a lesser extent than did holistic
score. They were: mean length of T-units (0.5504), error-free T-units
(0.5292), mean length of error-free T-units (0.4710), and overall
Table 2
Discriminant Analysis on Course Level Effects

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.2664</td>
<td>DV1</td>
<td>0.1816</td>
<td>0.4686</td>
<td>0.7827</td>
<td>: 0.4208</td>
</tr>
<tr>
<td></td>
<td>(0.9630)</td>
<td>DV2</td>
<td>-0.0269</td>
<td>-0.1952</td>
<td>0.1981</td>
<td>: F(12,194)=</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DV3</td>
<td>0.2390</td>
<td>0.8687</td>
<td>0.5504</td>
<td>: 8.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DV4</td>
<td>0.2248</td>
<td>0.9381</td>
<td>0.5292</td>
<td>: P &lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DV5</td>
<td>-0.0360</td>
<td>-0.1551</td>
<td>0.4710</td>
<td>:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DV6</td>
<td>0.0007</td>
<td>0.0694</td>
<td>0.5159</td>
<td>:</td>
</tr>
</tbody>
</table>

| 2        | 0.0487        | DV1      | -0.3052   | -0.7876     | -0.4772         | : 0.9536                                  |
|          | (0.0370)      | DV2      | 0.0038    | 0.0272      | 0.1264          | : F(5,98)=                                 |
|          |               | DV3      | 0.1775    | 0.6450      | 0.3823          | : 0.95                                    |
|          |               | DV4      | -0.0081   | -0.0339     | -0.2462         | : P < 0.4498                               |
|          |               | DV5      | -0.0497   | -0.2140     | -0.0624         | :                                        |
|          |               | DV6      | 0.0061    | 0.5790      | 0.3877          | :                                        |

PDP=Proportion of Discriminatory Power
DV1=Holistic Score; DV2=Number of T-units;
DV3=Mean Length of T-units; DV4=Number of Error-free T-units;
DV5=Mean Length of Error-free T-units; DV6=Overall Length
length (0.5159). The only variable that did not contribute to the discrimination of groups according to course level was the number of T-units.

The DA for the main effects for prompt type detected in the MANOVA resulted in one discriminant function that accounted for 100% of the total discriminatory power. As indicated in Table 3, the major variable contributing to the discrimination of groups according to topic was the holistic score (0.8688). The mean length of T-units (0.5928), mean length of error-free T-units (0.3824), and overall length (0.4746) also contributed to the discrimination of groups according to topic, but to much lesser degrees than did holistic score. In contrast, the structure coefficients representing the contribution of the number of T-units and of error-free T-units to the discrimination of groups according to topic were not significant.

The results of the DAs provide an examination of the contribution of the dependent variables to the discrimination of groups when the dependent variables are analyzed simultaneously. These results, however, are somewhat vague. In order to provide a thorough description of the detected differences, it is necessary to examine the variables individually because different degrees of contribution to the overall effects are revealed. This individual examination of variables was accomplished by means of separate univariate analyses and post-hoc multiple comparisons.
Table 3
Discriminant Analysis on Topic Effects

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6681</td>
<td>DV1</td>
<td>0.4234</td>
<td>1.0925</td>
<td>0.8688</td>
<td>0.5995</td>
</tr>
<tr>
<td></td>
<td>(1.0000)</td>
<td>DV2</td>
<td>-0.2181</td>
<td>-1.5816</td>
<td>0.0158</td>
<td>F(6,97)=</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DV3</td>
<td>-0.1276</td>
<td>-0.4638</td>
<td>0.5928</td>
<td>10.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DV4</td>
<td>-0.0569</td>
<td>-0.2375</td>
<td>-0.0220</td>
<td>P &lt; 0.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DV5</td>
<td>-0.0336</td>
<td>-0.1450</td>
<td>0.3824</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DV6</td>
<td>0.0179</td>
<td>1.6958</td>
<td>0.4746</td>
<td></td>
</tr>
</tbody>
</table>

PDP=Proportion of Discriminatory Power
DV1=Holistic Score; DV2=Number of T-units;
DV3=Mean Length of T-units; DV4=Number of Error-free T-units;
DV5=Mean Length of Error-free T-units; DV6=Overall Length
Results of the Univariate Analyses of Variance

Hypothesis 1A: There will be no significant difference among the three course levels as measured by holistic score. Table 4 shows that the average holistic score for subjects at the beginning level of ESL composition was 4.86. Subjects at the intermediate and advanced levels achieved mean holistic scores of 6.83 and 7.74, respectively. Table 5 shows that the ANOVA revealed highly significant differences among groups and thus led to a rejection of the null hypothesis, $F(2, 102) = 28.81, p < 0.0001$. Further main effect analysis using Scheffé's test of significant differences indicated that the intermediate and advanced groups were not significantly different from one another, but that both were significantly different from the beginning group.

Hypothesis 1B: There will be no significant difference between the general prompt and the field-specific prompt as measured by holistic score. An examination of Table 4 shows that the average holistic score was 5.20 for the general prompt and 7.17 for the field-specific prompt. The ANOVA resulted in highly significant differences between topics, $F(1, 102) = 48.12, p < 0.0001$, which thus led to a rejection of the null hypothesis (see Table 5). The post-hoc analysis conducted with Scheffé's test also indicated that the two topics were significantly different in terms of holistic score.
### Table 4
Means and Standard Deviations of Holistic Score as a Function of Course Level and Prompt Type

<table>
<thead>
<tr>
<th></th>
<th>General Prompt</th>
<th>Field Prompt</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Beginning (n=43)</td>
<td>4.21</td>
<td>1.68</td>
<td>5.51</td>
</tr>
<tr>
<td>Intermediate (n=43)</td>
<td>5.58</td>
<td>2.06</td>
<td>8.07</td>
</tr>
<tr>
<td>Advanced (n=19)</td>
<td>6.58</td>
<td>2.17</td>
<td>8.89</td>
</tr>
<tr>
<td>Overall</td>
<td>5.20</td>
<td>2.12</td>
<td>7.17</td>
</tr>
</tbody>
</table>

N=210

### Table 5
Analysis of Variance of Holistic Score by Course Level and Topic

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Ss</td>
<td>104</td>
<td>769.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Level (A)</td>
<td>2</td>
<td>277.68</td>
<td>138.84</td>
<td>28.81</td>
<td>0.0001</td>
</tr>
<tr>
<td>S(A)</td>
<td>102</td>
<td>491.58</td>
<td>4.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Ss</td>
<td>105</td>
<td>608.10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic (B)</td>
<td>1</td>
<td>189.64</td>
<td>189.64</td>
<td>48.12</td>
<td>0.0001</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>16.50</td>
<td>8.25</td>
<td>2.09</td>
<td>0.1286</td>
</tr>
<tr>
<td>SB/A</td>
<td>102</td>
<td>401.96</td>
<td>3.94</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>209</td>
<td>1377.36</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 1C: There will be no significant interaction between the course level variable and the prompt variable as measured by holistic score. The null hypothesis of no significant interaction between the two independent variables was retained for holistic score, $F(2, 102) = 2.09, p > 0.1286$.

Hypothesis 2A: There will be no significant difference among the three course levels as measured by number of T-units. Table 6 illustrates that subjects at the beginning level produced an average of 17.62 T-units per essay and subjects in the intermediate and advanced groups produced an average of 19.05 and 21.13 T-units per essay, respectively. The ANOVA (see Table 7) failed to detect a significant difference among the three groups, $F(2, 102) = 2.06, p > 0.1323$, thus leading to a retention of the null hypothesis.

Hypothesis 2B: There will be no significant difference between the general prompt and the field-specific prompt as measured by number of T-units. Table 6 shows an average number of T-units of 19.24 for the general prompt and of 18.44 for the field-specific prompt. This difference was not deemed significant by the ANOVA, $F(1, 102) = 3.24, p > 0.0749$ (see Table 7). The null hypothesis of no difference between topics was, therefore, retained.
Table 6
Means and Standard Deviations of T-units as a Function of Course Level and Prompt Type

<table>
<thead>
<tr>
<th></th>
<th>General Prompt</th>
<th>Field Prompt</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td>Beginning</td>
<td>17.40</td>
<td>5.63</td>
<td>17.84</td>
</tr>
<tr>
<td>(n=43)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intermediate</td>
<td>19.49</td>
<td>6.13</td>
<td>18.60</td>
</tr>
<tr>
<td>(n=43)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>22.84</td>
<td>13.00</td>
<td>19.42</td>
</tr>
<tr>
<td>(n=19)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>19.24</td>
<td>7.81</td>
<td>18.44</td>
</tr>
</tbody>
</table>

N=210

Table 7
Analysis of Variance of T-units by Course Level and Topic

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Ss</td>
<td>104</td>
<td>8539.49</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Level (A)</td>
<td>2</td>
<td>332.00</td>
<td>166.00</td>
<td>2.06</td>
<td>0.1323</td>
</tr>
<tr>
<td>S(A)</td>
<td>102</td>
<td>8207.49</td>
<td>80.47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Ss</td>
<td>105</td>
<td>2495.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic (B)</td>
<td>1</td>
<td>73.74</td>
<td>73.74</td>
<td>3.24</td>
<td>0.0749</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>98.57</td>
<td>49.29</td>
<td>2.16</td>
<td>0.1201</td>
</tr>
<tr>
<td>SB/A</td>
<td>102</td>
<td>2322.83</td>
<td>22.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>209</td>
<td>11034.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 2C: There will be no significant interaction between the course level variable and the prompt variable as measured by number of T-units. The null hypothesis of no interaction between the two independent variables was retained for the number of T-units, $F (2, 102) = 2.16, p > 0.1201$ (see Table 7).

Hypothesis 3A: There will be no significant difference among the three course levels as measured by mean length of T-units. Table 8 reveals that subjects in the beginning group produced T-units having a mean length of 13.19 words. The T-units produced by subjects at the intermediate and advanced levels had mean lengths of 14.51 words and 16.58 words, respectively. The ANOVA detected significant differences among course levels, $F (2, 102) = 10.71, p < 0.0001$, thus rejecting the null hypothesis of no difference (see Table 9). Post-hoc comparisons using Scheffé's test showed that the beginning and intermediate groups were not significantly different from one another, but that both were significantly different from the advanced group.

Hypothesis 3B: There will be no significant difference between the general prompt and the field-specific prompt as measured by mean length of T-units. The ANOVA for the prompt variable indicated a highly significant main effect for topic, $F (1, 102) = 15.81, p < 0.0001$ and thus led to a rejection of the null hypothesis (see Table 9). Table 8 reveals that when subjects responded in writing to the general prompt, they produced T-units having a mean length of 13.54 words. When the same subjects responded to the field-specific prompt, they produced
Table 8
Means and Standard Deviations of Mean Length of T-units as a Function of Course Level and Prompt Type

<table>
<thead>
<tr>
<th></th>
<th>General Prompt</th>
<th></th>
<th>Field Prompt</th>
<th></th>
<th>Overall</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>Beginning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=43)</td>
<td>12.53</td>
<td>2.62</td>
<td>13.85</td>
<td>3.04</td>
<td>13.19</td>
<td>2.90</td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=43)</td>
<td>13.70</td>
<td>4.20</td>
<td>15.32</td>
<td>3.15</td>
<td>14.51</td>
<td>3.78</td>
</tr>
<tr>
<td>Advanced</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=19)</td>
<td>15.50</td>
<td>4.18</td>
<td>17.66</td>
<td>3.04</td>
<td>16.58</td>
<td>3.77</td>
</tr>
<tr>
<td>Overall</td>
<td>13.54</td>
<td>3.75</td>
<td>15.14</td>
<td>3.35</td>
<td>14.34</td>
<td>3.63</td>
</tr>
</tbody>
</table>

N=210

Table 9
Analysis of Variance of Mean Length of T-units by Course Level and Topic

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Ss</td>
<td>104</td>
<td>1767.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Level (A)</td>
<td>2</td>
<td>306.91</td>
<td>153.46</td>
<td>10.71</td>
<td>0.0001</td>
</tr>
<tr>
<td>S(A)</td>
<td>102</td>
<td>1460.97</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Ss</td>
<td>105</td>
<td>991.80</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic (B)</td>
<td>1</td>
<td>132.48</td>
<td>132.48</td>
<td>15.81</td>
<td>0.0001</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>4.67</td>
<td>2.34</td>
<td>0.28</td>
<td>0.7572</td>
</tr>
<tr>
<td>SB/A</td>
<td>102</td>
<td>854.65</td>
<td>8.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>209</td>
<td>2759.68</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
significantly longer T-units (15.14 words). Post-hoc testing revealed a significant difference between the two topics as measured by mean length of T-units.

Hypothesis 3C: There will be no significant interaction between the course level variable and the prompt variable as measured by mean length of T-units. The null hypothesis of no significant interaction between the two independent variables was retained for mean length of T-units, \( F(2, 102) = 0.28, \ p > 0.7572 \).

Hypothesis 4A: There will be no significant difference among the three course levels as measured by the number of error-free T-units. Table 10 shows that subjects at the beginning level produced an average of 4.50 error-free T-units per essay. Subjects in the intermediate group produced an average of 7.15 error-free T-units and subjects in the advanced group produced an average of 8.47 error-free T-units per essay. A significant difference among groups was detected in the ANOVA for error-free T-units, \( F(2, 102) = 13.02, \ p < 0.0001 \), which resulted in a rejection of the null hypothesis (see Table 11). Post-hoc testing using Scheffé's test of significant differences revealed that the advanced and intermediate groups were not significantly different from one another. These two groups were, however, significantly different from the beginning group in terms of the overall number of error-free T-units.
Table 10
Means and Standard Deviations of Error-Free T-units as a Function of Course Level and Prompt Type

<table>
<thead>
<tr>
<th></th>
<th>General Prompt</th>
<th>Field Prompt</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Beginning</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=43)</td>
<td>4.86</td>
<td>3.75</td>
<td>4.14</td>
</tr>
<tr>
<td><strong>Intermediate</strong></td>
<td>7.42</td>
<td>3.82</td>
<td>6.88</td>
</tr>
<tr>
<td>(n=43)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Advanced</strong></td>
<td>8.89</td>
<td>4.71</td>
<td>8.05</td>
</tr>
<tr>
<td>(n=19)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>6.64</td>
<td>4.23</td>
<td>5.97</td>
</tr>
<tr>
<td><strong>N=210</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11
Analysis of Variance of Error-Free T-units by Course Level and Topic

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Ss</td>
<td>104</td>
<td>2558.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Level (A)</td>
<td>2</td>
<td>520.49</td>
<td>260.25</td>
<td>13.02</td>
<td>0.0001</td>
</tr>
<tr>
<td>S(A)</td>
<td>102</td>
<td>2038.01</td>
<td>19.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Ss</td>
<td>105</td>
<td>1081.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic (B)</td>
<td>1</td>
<td>22.36</td>
<td>22.36</td>
<td>2.16</td>
<td>0.1452</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>0.73</td>
<td>0.37</td>
<td>0.04</td>
<td>0.9655</td>
</tr>
<tr>
<td>SB/A</td>
<td>102</td>
<td>1057.94</td>
<td>10.37</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>209</td>
<td>3639.53</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hypothesis 4B: There will be no significant difference between the general prompt and the field-specific prompt as measured by the number of error-free T-units. Table 10 indicates that the mean number of error-free T-units for the general prompt was 6.64 and 5.97 for the field-specific prompt. The ANOVA did not detect a significant difference between these two means, \( F(1, 102) = 2.16, p > 0.1452 \) (see Table 11). This result led to a retention of the null hypothesis.

Hypothesis 4C: There will be no significant interaction between the course level variable and the topic variable as measured by the number of error-free T-units. The ANOVA revealed no significant interaction between the two independent variables as measured by error-free T-units, \( F(2, 102) = 0.04, p > 0.9655 \). This result led to a retention of the null hypothesis of no interaction.

Hypothesis 5A: There will be no difference among the three course levels as measured by the mean length of error-free T-units. Table 12 shows that the mean length of error-free T-units produced by subjects in the beginning group was 9.45 words. The mean length of error-free T-units produced by subjects in the intermediate and advanced groups was 11.49 and 12.96 words, respectively. The ANOVA detected a highly significant main effect for course level, \( F(2, 102) = 10.30, p < 0.0001 \), thus leading to a rejection of the null hypothesis of no difference (see Table 13). Scheffé’s test of significant differences revealed that the beginning group was significantly different from the
### Table 12
Means and Standard Deviations of Mean Length of Error-Free T-units as a Function of Course Level and Prompt Type

<table>
<thead>
<tr>
<th></th>
<th>General Prompt</th>
<th>Field Prompt</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>Beginning</strong> (n=43)</td>
<td>9.54</td>
<td>3.59</td>
<td>9.37</td>
</tr>
<tr>
<td><strong>Intermediate</strong> (n=43)</td>
<td>11.16</td>
<td>5.32</td>
<td>11.83</td>
</tr>
<tr>
<td><strong>Advanced</strong> (n=19)</td>
<td>11.46</td>
<td>2.41</td>
<td>14.46</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>10.55</td>
<td>4.28</td>
<td>11.30</td>
</tr>
</tbody>
</table>

N=210

### Table 13
Analysis of Variance of Mean Length of Error-Free T-units by Course Level and Topic

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between Ss</strong></td>
<td>104</td>
<td>2208.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Level (A)</td>
<td>2</td>
<td>371.16</td>
<td>185.58</td>
<td>10.30</td>
<td>0.0001</td>
</tr>
<tr>
<td>S(A)</td>
<td>102</td>
<td>1837.03</td>
<td>18.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within Ss</strong></td>
<td>105</td>
<td>1703.44</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic (B)</td>
<td>1</td>
<td>61.09</td>
<td>61.09</td>
<td>3.95</td>
<td>0.0495</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>66.18</td>
<td>33.09</td>
<td>2.14</td>
<td>0.1227</td>
</tr>
<tr>
<td>SB/A</td>
<td>102</td>
<td>1576.17</td>
<td>15.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>209</td>
<td>3911.63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
intermediate and advanced groups, but that the latter two were not significantly different from one another.

Hypothesis 5B: **There will be no significant difference between the general prompt and the field-specific prompt as measured by the mean length of error-free T-units.** The means presented in Table 12 show that the overall mean length of T-units produced by subjects was 10.55 words for the general topic and 11.30 words for the field-specific topic. The ANOVA detected a significant difference between the two prompts at the .05 level of significance, \( F(1, 102) = 3.95, p < 0.0495 \), and led to a rejection of the null hypothesis. Post-hoc comparisons of the two means using Scheffé's test did not, however, result in significance.

Hypothesis 5C: **There will be no interaction between the course level variable and the prompt variable as measured by the mean length of error-free T-units.** The null hypothesis of no interaction between the two independent variables was retained for the mean length of error-free T-units, \( F(2, 102) = 2.14, p > 0.1227 \) (see Table 13).

Hypothesis 6A: **There will be no difference among the three course levels as measured by overall length.** Table 14 reveals that the mean number of words produced by subjects in the beginning group was 225.16. The intermediate group produced an average of 263.49 words per essay, and the advanced group produced an average of 326.24 words per essay. The ANOVA results reported in Table 15 revealed a highly significant
### Table 14
Means and Standard Deviations of Overall Length as a Function of Course Level and Prompt Type

<table>
<thead>
<tr>
<th></th>
<th>General Prompt</th>
<th></th>
<th>Field Prompt</th>
<th></th>
<th>Overall</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td><strong>Beginning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=43)</td>
<td>213.91</td>
<td>75.52</td>
<td>236.42</td>
<td>72.91</td>
<td>225.16</td>
<td>74.65</td>
</tr>
<tr>
<td><strong>Intermediate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=43)</td>
<td>251.65</td>
<td>64.23</td>
<td>275.33</td>
<td>79.47</td>
<td>263.49</td>
<td>72.81</td>
</tr>
<tr>
<td><strong>Advanced</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(n=19)</td>
<td>324.63</td>
<td>155.50</td>
<td>327.84</td>
<td>114.92</td>
<td>326.24</td>
<td>134.88</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>249.40</td>
<td>98.55</td>
<td>268.90</td>
<td>89.84</td>
<td>259.15</td>
<td>94.57</td>
</tr>
</tbody>
</table>

N=210

### Table 15
Analysis of Variance of Overall Length by Course Level and Topic

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>PR &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Ss</td>
<td>104</td>
<td>1611953.93</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Course Level (A)</td>
<td>2</td>
<td>271984.35</td>
<td>135992.18</td>
<td>10.35</td>
<td>0.0001</td>
</tr>
<tr>
<td>S(A)</td>
<td>102</td>
<td>1339969.58</td>
<td>13136.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Ss</td>
<td>105</td>
<td>250015.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Topic (B)</td>
<td>1</td>
<td>12629.94</td>
<td>12629.94</td>
<td>5.50</td>
<td>0.0210</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>3090.45</td>
<td>1545.23</td>
<td>0.67</td>
<td>0.5126</td>
</tr>
<tr>
<td>SB/A</td>
<td>102</td>
<td>234294.67</td>
<td>2297.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>209</td>
<td>1861988.99</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
difference among the course levels for overall length, $F(2, 102) = 10.35$, $p < 0.0001$, thus leading to a rejection of the null hypothesis of no difference. Scheffé's test of significant differences showed that the beginning and intermediate groups were not significantly different from one another, but that both were significantly different from the advanced group.

Hypothesis 6B: There will be no significant difference between the general prompt and the field-specific prompt as measured by overall length. The means for overall length according to prompt type are reported in Table 14. The mean number of words produced for the general prompt was 249.40 and 268.90 for the field-specific prompt. The ANOVA detected a significant difference between the two topics for overall length, $F(1, 102) = 5.50$, $p < 0.0210$, thus leading to a rejection of the null hypothesis of no difference (see Table 15). Post-hoc testing also revealed a significant difference between the two topics as measured by overall length.

Hypothesis 6C: There will be no interaction between the course level variable and the prompt variable as measured by overall length. The ANOVA for overall length did not detect a significant interaction between the two independent variables, $F(2, 102) = 0.67$, $p > 0.5126$. The null hypothesis of no interaction was, therefore, retained.
Discussion

The failure of the MANOVA to detect a significant interaction between the independent variables was anticipated because the groups were formed on the basis of ESL course level, and familiarity with topic was expected to have comparable influence on all groups. In other words, subjects' scores were expected in general to increase proportionately with course level.

The results of the DA on the main effect for course level indicate that subjects' writing performance improved according to course level on the basis of an increase in the scores of all dependent variables except the number of T-units. In other words, subjects at higher course levels wrote essays that received higher scores on all the variables except number of T-units. These results suggest that subjects in the intermediate and advanced groups made fewer errors, wrote longer essays, and produced more syntactically complex utterances (as indicated by the decrease in their number of T-units and simultaneous increase in the mean length of those T-units) than did subjects in the beginning group. These characteristics, in turn, probably led to the higher holistic scores received by subjects at the higher levels.

The results of the DA on topic effects reveal that an improvement in a subjects' writing performance on the basis of topic was due to an increase in the holistic score and overall length of the subjects' written response to the field-specific topic. An increase in syntactic complexity (as indicated by the increase in the mean length of T-units
and error-free T-units) also contributed to the discrimination of groups according to topic.

The results of the univariate ANOVAs on the individual dependent variables were, as expected, in direct agreement with the results of the DAs. The univariate results, however, serve to explain more clearly the discrimination among groups according to the course level and topic variables. In order to facilitate the discussion on the results of the univariate analyses, graphic representations of the group means on the dependent variables for both topics are presented.

The failure of the univariate ANOVA to detect a significant interaction between the independent variables for holistic score was not unexpected for two reasons. First, the MANOVA did not detect a significant interaction. Secondly, holistic scores were found in the DAs to be the major variable contributing to the discrimination of groups according to both course level and topic.

The results of the post-hoc test performed on the main effect for course level as measured by holistic score provide preliminary data in terms of the developmental aspects of different groups of ESL writers. As seen in Figure 2, holistic scores increased in accordance with course level. The difference in this increase between the advanced and intermediate groups, however, did not prove to be significant. Consequently, at least in terms of holistic score, these two groups are less distinguishable from one another than they are from the beginning group. This result suggests that in order for fine distinctions to be made between writing samples of subjects at higher levels, it may be
Figure 2: Main Effects of Course Level and Topic as Measured by Holistic Score
necessary to examine the writing samples on the basis of criteria in addition to holistic scores.

The graph in Figure 2 illustrates the marked increase in holistic scores for all groups on the field-specific topic. The result of significant main effects for topic supports findings from L1 studies that subjects' writing performance improves qualitatively when the subjects are familiar with the topic of the writing stimulus (Cheskey, 1984; Cheskey & Hiebert, 1987; Langer, 1984). As expected, L2 writers, like L1 writers, produce qualitatively better writing, as measured by holistic scores, when provided with a topic that allows them to make use of their prior knowledge.

Although the ANOVA on the number of T-units did not result in a significant interaction between course level and topic, it should be noted that a slight ordinal interaction does exist between the variables, as illustrated in Figure 3. Subjects in the beginning group wrote slightly more T-units on the field-specific topic than they did on the general topic. In contrast, the intermediate and advanced groups produced more T-units on the general topic than on the field-specific topic. The failure of the ANOVA to detect a significant interaction may be due to the large amount of within-group variability in the advanced group. This high within-group variability led to an increase in the SB/A mean square error term, which, in turn, weakened the capability of the ANOVA to find a significant interaction.

The ANOVA for number of T-units also resulted in non-significant main effects. A close examination of the graph in Figure 3, however, prompted the researcher to investigate more thoroughly the difference
Figure 3: Main Effects of Course Level and Topic as measured by Mean Number of T-units
between topics in terms of the number of T-units produced by subjects in the advanced group. The minimum difference between the two means at the advanced level that is necessary to claim a significant difference between the topics was determined by means of Fisher's Least Significant Difference (LSD) formula. The critical value for the LSD was calculated at 3.065. The difference between the advanced group's mean number of T-units for the general prompt (22.84) and for the field-specific prompt (19.42) was 3.42. Consequently, the difference between the mean numbers of T-units produced by advanced subjects on the two topics is greater than would be expected if due to error. In other words, advanced subjects produced significantly fewer T-units on the field-specific topic than on the general topic. The underlying meaning of this difference in number of T-units is dependent on further examination of the mean length of those T-units.

The graph in Figure 4 clearly shows that the mean length of T-units increased in accordance with course level. In other words, advanced subjects produced longer T-units than intermediate subjects, who produced longer T-units than beginning subjects. The mean length of T-units serves as an indication of the different subjects' ability to produce more syntactically complex utterances. The result of Scheffé's test suggests that the beginning and intermediate groups were more similar in terms of their ability to produce utterances having greater syntactic complexity. Subjects in the advanced group were clearly more mature than subjects in the other two groups in terms of their syntactic development. This result parallels findings of L1 studies that have found mean length of T-units to be valid indicators of
Figure 4: Main Effects of Course Level and Topic as Measured by Mean Length of T-units
syntactic development. It must be noted, however, that error-free T-units have been found to be more valid indices of syntactic development in L2 learners than the basic T-unit index. For this reason, it is necessary to examine the results of the analyses of the T-unit indices in conjunction with the results of the analyses of the error-free T-unit indices in order to arrive at an accurate interpretation of L2 learners' ability to produce utterances having increasing degrees of syntactic complexity.

As the graph in Figure 4 clearly shows, all subjects regardless of course level produced more syntactically complex utterances on the topic with which they were familiar. For this reason, the advanced group produced significantly fewer T-units on the field-specific prompt than on the general prompt. In other words, because the advanced subjects produced longer T-units for the field-specific topic, they produced fewer T-units overall on that topic.

The significant difference between topics as measured by mean length of T-units that resulted in this analysis was not consistent with the results on mean T-unit length reported in Cheskey's (1984) study with L1 writers. Cheskey (1984; Cheskey & Hiebert, 1987) found no significant difference between high- and low-prior knowledge groups on T-unit length. This inconsistency of findings may be due to the nature of the topics themselves. Several researchers have found that different purposes in writing tasks result in differing degrees of syntactic complexity in the written responses (Crowhurst & Piché, 1979; Quellmalz et al., 1982; Rubin & Piché, 1979). In addition, the differences between L1 and L2 writers may be another reason for the
inconsistency in these results. L2 writers may demonstrate more variability than L1 writers in terms of the syntactic complexity of their writing, depending on the topic of the writing stimulus.

As expected, subjects at the beginning level made more errors than subjects at the other two levels and, consequently, produced the smallest number of error-free T-units across the three groups. The advanced and intermediate groups were similar in terms of the number of error-free T-units they produced.

The graph in Figure 5 clearly illustrates that all three groups produced fewer error-free T-units in the essays they wrote on the topic with which they were familiar. This result, like the similar result reported earlier for the number of T-units, is difficult to interpret without examining the results of the ANOVA for the mean length of these error-free T-units.

Although the ANOVA on mean length of error-free T-units did not result in a significant interaction between course level and topic, as the graph in Figure 6 shows, a slight ordinal interaction did occur between course level and prompt type. It may be that the ANOVA was unable to find a significant interaction because of the large value of the SB/A mean square error term.

The results of Scheffé's test concerning the main effect for course level as measured by the mean length of error-free T-units must be examined in conjunction with the result reported for the number of error-free T-units. The beginning group is clearly distinguished from the other two groups on the basis of the error-free T-unit indices. The distinction between the intermediate and advanced groups is not,
Figure 5: Main Effects of Course Level and Topic as Measured by Mean Number of Error-Free T-Units
Figure 6: Main Effects of Course Level and Topic as Measured by Mean Length of Error-Free T-Units
however, so clearly defined in terms of the error-free T-unit indices. A close examination of Figure 6 suggests, however, that a distinction between the intermediate and advanced groups as measured by the mean length of error-free T-units may be more clearly defined on the basis of topic differences.

Post-hoc testing of the significant main effect for topic as measured by mean length of error-free T-units did not result in significance. A close examination of the graph in Figure 6, however, prompted the researcher to investigate more thoroughly the difference between the two topics on the basis of the mean length of error-free T-units produced by the advanced group. As Figure 6 clearly illustrates, the intermediate and advanced groups performed similarly on the general topic as measured by the mean length of error-free T-units. There was very little difference between their respective means of 11.15 and 11.46 words per error-free T-unit for the general topic.

In contrast, the performance of these two groups on the field-specific topic differed greatly in terms of their mean length of error-free T-units. The intermediate group showed a slight increase in the number of words per error-free T-unit (11.83) on the field-specific topic, whereas the advanced group displayed a marked increase in the number of words they produced per error-free T-unit (14.46) on the field-specific topic. Fisher’s LSD was calculated in order to determine whether the main effect of prompt type would be significant for the advanced group. The critical value calculated for the LSD was 2.525, which represents the minimum difference between the means that
is necessary in order to claim significance at the .05 level. The
difference between the means of the advanced group was 3.00.
Consequently, the advanced group did, in fact, produce significantly
longer error-free T-units on the field-specific topic than they did on
the general topic.

This finding is noteworthy for several reasons. Familiarity with
topic appears to influence subjects at the advanced level more than
subjects at the other levels in terms of syntactic complexity. The
result of the analysis on the mean length of T-units, which was
reported earlier, was consistent with this finding. These results
suggest that when L2 writers are mature in terms of linguistic
competence, they are able to demonstrate the level of their competence
when given the opportunity to respond in writing to a topic with which
they are familiar. Conversely, when L2 writers' linguistic competence
is lacking, their familiarity with the topic of the writing stimulus
does not compensate for their lack of linguistic competence.
McCutcheon's (1986) study with L1 children resulted in similar
findings. Younger children's extensive content knowledge on the topic
of the writing task did not compensate for their immature writing
strategies. The findings of this analysis suggest that, on the one
hand, if L2 writers have a limited amount of linguistic knowledge in
the L2, their familiarity with the topic of a writing task does not
provide them with the linguistic knowledge also required to produce
quality writing. If, on the other hand, L2 writers are capable of
producing syntactically complex utterances with fewer errors, their
familiarity with the topic allows them to demonstrate this capability.
The findings in this analysis also suggest a possible reason why the advanced group produced fewer error-free T-units overall on the field-specific topic than they did on the general topic. Subjects at the advanced level produced more syntactically complex utterances for the field-specific prompt, as indicated by the mean length of T-units and error-free T-units. It may be posited that these syntactically complex utterances created more possibilities for error to occur, which, in turn, resulted in a smaller number of error-free T-units overall.

As reported previously, post-hoc testing among the three course levels for the mean length of error-free T-units did not detect a significant difference between the means of the intermediate and advanced groups. The findings in this analysis, which are clearly illustrated in Figure 6, suggest, however, that the subjects' writing performance on the field-specific topic leads to a clearer distinction between these two groups than their performance on the general topic.

The significant main effect for course level found in the ANOVA on overall length lends support to earlier findings that overall length of written text tends to increase proportionally with levels of L2 proficiency (Larsen-Freeman & Strom, 1977; Neuman, 1977). Similarly, the main effect for topic as measured by overall length parallels claims of studies with L1 writers that subjects tend to produce longer texts on topics with which they are familiar (Cheskey, 1984; Cheskey & Hiebert, 1987; McCutcheon, 1986). An L2 study conducted by Winfield and Barnes-Felfeli (1982) also indicated that subjects who were
familiar with the topic of the writing stimulus produced more words on that topic than on the topic with which they were not familiar.

Figure 7 illustrates the similarity between the beginning and intermediate groups' performance on the two topics in terms of overall length, as indicated by the comparable slope in the graph. The increase in the overall length for the field-specific topic produced by subjects at both the beginning and intermediate levels was approximately the same. This finding suggests that beginning and intermediate subjects demonstrate more fluency with the language when responding in writing to a topic with which they are familiar than when responding to a general topic.

In contrast, topic did not appear to make much difference in the number of words produced by the advanced subjects. This result suggests that subjects at the advanced level had achieved a degree of linguistic competence that allowed them to demonstrate fluency with the language regardless of their familiarity with the topic of the writing stimulus.
Figure 7: Main Effects of Course Level and Topic as Measured by Overall Length
CHAPTER V
SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Overview of the Study

The present study has applied the theoretical knowledge that has emerged from research on cognition, and L1 and L2 reading and writing to the assessment of writing in ESL learners. The purpose of this study was to add to the exceedingly limited amount of L1 and L2 writing research concerning the effects of topic variables on writing performance and to examine these effects under assessment conditions. The study investigated the extent to which ESL graduate students’ writing performance was affected by their knowledge of the subject matter of the topic used to elicit their writing performance.

Graduate students at three different course levels of ESL composition responded to a topic of general nature and to a topic pertaining to their fields of study. Their writing performance was assessed for overall quality, syntactic complexity, and fluency. These characteristics were measured by holistic ratings, counts of T-unit and error-free T-unit indices, and overall length. The data were subjected to several statistical analyses, including a multivariate analysis of variance, discriminant analyses, and univariate analyses of variance followed by post-hoc multiple comparisons.
Summary of Findings and Conclusions

Overview

The data generated by this study support findings reported in studies with L1 writers (Cheskey, 1984; Langer, 1984) that prior knowledge of the topic of a writing stimulus improves the quality of writing performance. An analysis of the data demonstrated that the general topic was not significantly superior to the field-specific topic on any of the writing measures examined. Furthermore, the data revealed that the field-specific topic resulted in statistically significant increases on three of the six writing measures examined. The field-specific topic also resulted in more consistently proportional increases across the three course levels than the general topic on the majority of the writing measures. The overall findings suggest that topics that are familiar to writers elicit more accurate representations of the writers' underlying proficiency than do general topics.

The multivariate analysis of variance (MANOVA) detected no significant interaction between the course level and topic variables when all six dependent variables were analyzed simultaneously. This result was anticipated because the groups were formed on the basis of their ESL course level, and familiarity with the topic of the writing stimulus was expected in general to influence all groups. That is, subjects' scores were expected to increase proportionally with course levels. The subsequent univariate analyses of variance (ANOVAs) detected no significant interaction either, which was not surprising in light of the fact that no interaction occurred in the MANOVA.
Highly significant main effects for both the course level variable and the topic variable were detected in the MANOVA. The discriminant analyses (DAs) that were performed subsequent to the MANOVA indicated that all of the dependent variables made significant contributions to the discrimination of groups according to course level except for the number of T-units. Of the significant variables, holistic score was found to be the major variable contributing to the discrimination of groups according to course level. Similarly, the DA performed on the main effects for topic found holistic score to be the major variable contributing to the discrimination of groups according to topic. The variables related to length and to syntactic complexity (i.e., mean length of T-units and of error-free T-units, and overall length) were also found to contribute significantly to the discrimination among groups according to topic. The influence of these variables as described in the DAs was explained more fully in the univariate ANOVAs.

Main Effects for Course Level

The findings of this study that are related to the main effects for course level provide information concerning the developmental characteristics of ESL learners at different course levels. The beginning group was clearly distinguished from the intermediate and advanced groups on the basis of holistic score and the error-free T-unit indices. Because holistic scores were determined on the basis of a variety of qualitative features, such as overall expression, organization, coherence, and syntactic variety and accuracy, it may be posited that subjects at the beginning level were clearly inferior to subjects at the other levels in terms of their knowledge of English
text structure and syntax, as evidenced by their lower holistic ratings. Purves and Purves (1986) claim that writers need to tap several kinds of knowledge in the activity of writing: linguistic knowledge, discourse knowledge, knowledge of pragmatics, and subject matter knowledge. The beginning subjects in this study appear to be developmentally inferior to the intermediate and advanced subjects in terms of their linguistic and discourse knowledge bases. This assertion is also supported by the fact that the beginning subjects made more syntactic and lexical errors, as evidenced by their production of significantly fewer error-free T-units, which were also significantly shorter than those produced by subjects at the intermediate and advanced levels.

In contrast, the writing performance of beginning subjects was not significantly different from that of the intermediate and advanced subjects in terms of the number of T-units produced, as was also indicated in the DA. This finding is meaningless, however, if discussed apart from the results concerning the mean length of those T-units and overall length because these three variables are interrelated in terms of their calculations. The advanced group was significantly different from the other groups on the basis of mean length of T-units and overall length. This finding demonstrates the superior syntactic development of the advanced subjects, as evidenced by the mean length of T-units, and their superior fluency with the language, as evidenced by overall length. These subjects were clearly more capable than the other subjects to write longer and more syntactically complex texts in the limited amount of time allotted.
The intermediate group was the least definable of the three groups in terms of developmental aspects. Post-hoc comparisons revealed that subjects at the intermediate level were found to be similar to beginning subjects with regard to their fluency with the language. In contrast, they were found to be similar to advanced subjects on measures of holistic scores and error-free T-unit indices. The similarities between the intermediate and advanced subjects, however, must be interpreted with caution.

The failure of the post-hoc comparisons to detect significant differences between the intermediate and advanced groups may be due to the smaller number of subjects in the advanced group. The greater variability in the intermediate group may have made the existence of significant differences between the groups difficult to detect. As explained previously, the advanced and intermediate groups produced error-free T-units of approximately the same mean length on the general topic. On the field-specific topic, however, the intermediate group's mean showed only a slight increase, whereas the advanced group's mean showed a sharp increase. This finding suggests, then, that the subject matter of the writing stimulus may be a crucial factor in arriving at clear distinctions between adjacent levels of writers.

Several researchers have claimed that error-free T-unit indices are not often sensitive enough to discriminate among learners at adjacent levels of L2 proficiency (Cooper, 1976; Gaies, 1980; Larsen-Freeman & Strom, 1976; Monroe, 1975). The results of this study suggest that the inconsistency of the capability of error-free T-unit indices to detect differences between learners at adjacent levels may be due to
differences in the topics used to elicit the writing performance rather than to problems inherent in the measurement index itself. In other words, it may be argued that general topics that do not allow writers to make use of their prior knowledge do not allow writers to demonstrate their underlying ability to write in the L2. The findings in this study concerning error-free T-unit indices lend strong support to this claim. Other findings reported in L1 studies also support this assertion. For example, several researchers have found that aspects of topic variables, such as mode of discourse, affects the degree of syntactic complexity that is produced in writing (Crowhurst & Piché, 1979; Quellmalz et al., 1982; Rubin & Piché, 1979).

In sum, the findings of the present study concerning main effects for the group variable have demonstrated developmental aspects of ESL writers at different course levels. On the whole, as subjects' course level increased, the quality of their writing improved, as indicated by the increase in holistic scores and by the proportional increase in the mean length of T-units and error-free T-units for the field-specific topic. In addition, the results of this study are consistent with the findings reported in other L2 studies on writing that overall length and error-free T-unit indices appear to be indicators of growth in L2 proficiency (e.g., Gaies, 1980; Larsen-Freeman & Strom, 1977; Neuman, 1977).

Main Effects for Topic

The main effects detected for the variable of prompt type revealed that the field-specific topic was clearly superior to the general topic as measured by holistic score, mean length of T-units, and overall
length. That is, when the L2 writers in this study were given the opportunity to respond to a topic with which they were familiar, they produced longer essays that were qualitatively better than the essays they produced in response to the general topic in the limited amount of time allotted. These findings suggest that because the subjects had prior knowledge about the field-specific topic, they had previously thought about the topic and perhaps had read about it and discussed it with others. It may be posited, then, that the subjects had a good idea about how aspects of their knowledge on the topic fit together to form a coherent, organized whole. In other words, it is likely that their holistic scores increased on the field-specific topic because their knowledge base was organized, contextualized, and more fluent.

In contrast, it may be hypothesized that the subjects in this study had never previously given much thought to the subject matter of the general prompt. Consequently, when faced with a limited amount of time in which to compose a response, the subjects may have found it overwhelming to formulate ideas about the topic and to organize them into a unified piece of academic discourse. This hypothesis is supported by several of the findings on the topic variable.

Subjects in all groups received higher holistic ratings on the field-specific topic than on the general topic. In addition, the beginning and intermediate groups produced significantly longer texts and longer T-units on the field-specific topic than on the general topic. Although the length of the texts produced by the advanced subjects remained constant over both topics, significantly fewer, but significantly longer T-units were produced by the advanced subjects
on the field-specific topic. This finding illustrates that the advanced subjects' fluency in English was such that they were able to write approximately the same amount of text for both prompts, but that their syntactic complexity was superior for the field-specific prompt. In other words, the advanced subjects' texts were quantitatively similar, but qualitatively very different.

The superior quality of the advanced subjects' writing on the field-specific topic was also evidenced by the results of the error-free T-unit measures. Although post-hoc comparisons did not indicate a significant difference between the topics as measured by mean length of error-free T-units, a calculation of Fisher's Least Significant Difference formula showed that whereas the intermediate subjects did not produce significantly longer error-free T-units on the field-specific topic than they did on the general topic, the advanced subjects did. Consequently, familiarity with the topic clearly enabled the advanced subjects to demonstrate their superior proficiency with the language.

In contrast, although they were able to make use of their prior knowledge, the beginning subjects were not capable of producing longer error-free T-units on the field-specific topic than on the general topic. As mentioned earlier, this finding mirrors McCutcheon's (1986) finding in children writing in their L1: prior knowledge of the subject matter of the topic does not compensate for immature writing strategies. Another finding reported by McCutcheon (1986), however, is debated by the findings of this study. McCutcheon (1986) claimed that mature writing strategies, on the other hand, could compensate for lack
of prior knowledge on the topic. If her claim were true, the advanced
subjects in the present study should have performed similarly on both
topics in terms of their holistic ratings and indices of syntactic
complexity. This was not the case. On the contrary, only when
advanced subjects were able to tap their subject matter knowledge were
they able to demonstrate their superior linguistic development. Again,
topic familiarity appeared to be of major influence on writing
performance.

Significant differences were not detected between the two topics as
measured by the number of error-free T-units. In fact, subjects at all
course levels wrote fewer error-free T-units on the field-specific
topic than on the general topic, though this difference was not deemed
as statistically significant. At first, this finding may be understood
to be contradictory to the superiority of the field-specific topic.
When this finding is viewed, however, in conjunction with the result
that revealed a significant difference between the topics as measured
by the mean length of T-units, a different understanding occurs.
Subjects at all course levels wrote significantly longer T-units on the
field-specific topic. It may be posited, then, that these longer
T-units created more possibilities for error to occur, which, in turn,
resulted in the production of fewer error-free T-units overall on the
field-specific topic.

Yet another variable that was not directly observed in the present
study, but that is deserving of discussion, is the variable of interest
in the topic used to elicit the writing performance. Recall that the
subjects responded to the general prompt during the first phase of the
experiment. This phase took place during the actual administration of the diagnostic examination in the ESL composition courses for Fall Quarter, 1987. When the subjects wrote on the general topic, they knew that the results of their writing performance could potentially serve to move them to a higher level in the series of ESL courses, or to qualify them. In other words, the subjects were provided with external motivation for performing well on the general topic. In contrast, the subjects were informed prior to the administration of the field-specific writing exercise that the results of their writing performance on this task were for research purposes only and would in no way affect their status in the ESL program. Despite the apparent lack of external motivation for the second writing task, the subjects' writing performance on this task showed significant improvement over their writing on the first task. It appears that the subjects may have created an internal motivation that resulted in an improvement in their performance. This motivation may have stemmed from the subjects' personal interest in the subject matter of the field-specific prompt.

In sum, the results of the present study clearly suggest that topic familiarity does indeed lead to an improvement in the writing performance of L2 writers. In addition, topics that allow writers to make use of their prior knowledge are superior to general topics in terms of their ability to discriminate among groups having different levels of writing proficiency. These findings carry important implications for ESL writing assessment and pedagogical practices.
Implications for Assessment and Pedagogy

Important decisions concerning L2 students' acceptance into universities and/or academic programs and their placement into various courses are often made on the basis of writing assessments. The findings of the present study clearly indicate that the extent to which L2 writers are familiar with the subject matter of the writing stimulus has dramatic influences on their writing performance. It is imperative that topics be developed for use in writing assessment that allow writers to make use of their prior knowledge, because topics that tap subject matter knowledge appear to elicit the kind of writing required to make accurate judgments about students' underlying writing ability.

The results of this study bring into question the validity of the current instruments used to assess L2 writing proficiency. Neither the ACTFL Provisional Proficiency Guidelines nor the TWE, which were both described earlier, contain writing stimuli that allow writers to make use of their prior knowledge. The present study provides evidence to suggest that these instruments do not elicit writing performance that accurately represents the writers' underlying writing proficiency. These instruments are in need of modification, particularly if they are to be used to make decisions regarding students' admission into universities and placement into academic programs.

The results of this study also provide practical considerations for ESL writing assessment. For example, in many ESL programs, diagnostic and placement examinations are administered for the purpose of discovering where individual writers fit into a series of courses and for the purpose of detecting individual writers' strengths and
weaknesses. Those instructors or directors in charge of making
decisions on the basis of the writing samples are usually interested in
devising topics that will elicit lengthy samples of writing. In
addition, they are interested in obtaining samples that demonstrate
syntactic and lexical variety and that provide information about the
writers' ability to produce syntactically complex utterances and to
create organized, coherent discourse. The results of this study show
that topics that allow subjects to utilize their prior knowledge meet
the above criteria. The field-specific topic elicited longer responses
overall. In addition, this topic appeared to encourage students to take
risks in terms of producing syntactically complex utterances, as
evidenced by the longer T-units produced by all groups, which, in turn,
appeared to result in fewer error-free T-units.

In relation to pedagogical practices, the findings of the present
study demonstrate that differences in topic affect writing performance
in critical ways. Topics that are assigned in ESL classrooms should
allow subjects to demonstrate their underlying writing ability. In
addition, topics should encourage risk-taking on the part of writers.
That is, writers should be encouraged to attempt to produce
increasingly complex utterances. The results of the relationship
between subject matter knowledge and writing performance as measured by
holistic scores in the present study also suggest that attention should
be given in the classroom to prewriting and brainstorming sessions so
that students can see the importance of utilizing and organizing what
they know about a topic. This suggestion also implies that instructors
need to take the time to discover what their students know so that they
can help their students to grow as writers. Instructors would also be wise to have students generate topics for classroom assignments and examinations.

This study also implies that interest in the topic of a writing task on the part of the students influences their writing performance. In his study with L1 writers, Cheskey (1984) found interest to be significantly correlated with prior knowledge of topic. ESL instructors and test developers should actively pursue topics that engage the interest of ESL writers.

Recommendations for Further Research

Writing is such a complex phenomenon that it would be short-sighted to believe that quantifiable variables, such as those examined in the present study, provide researchers with all of the answers about the activity of writing. Quantitative variables are valuable in that they reveal patterns among subjects and pinpoint areas that would benefit from indepth qualitative analyses. What is clearly needed are qualitative data concerning the differences in writing performance as influenced by topic familiarity. For example, high holistic scores suggest that essays are qualitatively better than essays with low holistic scores. But these scores do not define quality. It is, therefore, necessary to examine writing samples, such as those collected in this study, qualitatively, in order to arrive at a clearer understanding of the factors that distinguish between essays receiving high holistic ratings and essays receiving low holistic ratings.
Similarly, larger measures of T-units and error-free T-units suggest greater syntactic complexity in the writing. This syntactic complexity, however, warrants closer examination. Research needs to determine exactly which factors account for increases in T-unit and error-free T-unit indices. For example, are longer error-free T-units the result of the addition of embedded relative clauses, or are they simply the result of compound nominal clauses?

Because this study is the first of its kind involving the influence of topic familiarity on ESL learners' writing performance in an assessment setting, replications and enhanced experiments are certainly in order. This study has focused on graduate ESL students. Other studies involving undergraduates and younger ESL students are necessary, particularly in light of the growing numbers of ESL students in the United States at all levels of academic study, and of the growing emphasis on writing assessment. In addition, studies with American foreign language (FL) learners should be conducted. Most of the L2 literature base in writing stems from research conducted with ESL learners. In view of the recent interest in FL study in the United States, and the recent focus on the testing of FL proficiency, studies of this kind with FL learners are clearly needed.

This study has focused on only one aspect involving topic variables. In addition to further studies on subject matter influences, similar experiments involving other topic variables, such as wording, rhetorical specification, and mode of discourse, are required. These studies may provide the needed information about the different kinds of knowledge (e.g., discourse knowledge, knowledge of
pragmatics, etc.) discussed by Purves and Purves (1986). In addition, attention must be focused on how different topics are perceived by the different individuals involved in a writing assessment episode. Ruth and Murphy (1984, 1988) have suggested that topic variables may lead to different interpretations of the same topic on the part of the test maker, the test taker, and the test rater. These different interpretations may dramatically influence the assessment of an individual's writing performance.

In view of the present study's indirect implication that interest appears to be related to writing performance in addition to topic familiarity, experiments are needed that investigate the degree of writers' interest in and commitment to certain topics. Information related to writers' interest levels may be gathered by means of involvement surveys, such as the one used by Cheskey (1984), or by means of self-reports, such as those often used in research studies on the writing process.

Finally, studies comparing the effects of topic variables on the writing performance of both L1 and L2 writers is needed. At the university level, international students attend classes with American students and are usually expected to perform comparably to Americans on assignments and examinations related to the content of the course. Many of the examinations administered in graduate level courses in particular involve essay questions. International students may be at a disadvantage with these examinations not because they lack an understanding of the content of the course, but because they misinterpret the point of the question and respond inappropriately.
For example, native and nonnative writers of English may interpret writing prompts differently depending upon their respective degrees of sociocultural understanding in terms of the wording of the topics. Studies are needed to determine the extent to which L1 and L2 learners interpret topics differently and to discover how those interpretations influence their responses.

In conclusion, research on topic variables is in its infancy. The possibilities for empirical investigation in this area are innumerable. The importance of the research on topic variables as they relate to the assessment of writing cannot be underestimated in view of the important decisions that are made on the basis of writing assessment.

**Limitations of the Study**

Whenever collecting repeated measures from subjects, researchers are cautioned to control for effects to internal validity caused by the extraneous variable known as testing (Campbell & Stanley, 1963). The ideal way to control for testing effects involves randomly assigning the different levels of the treatment variable to subjects at the various times at which repeated measures are obtained. It was not possible, however, to exercise this control in the present study because the first measures were obtained during the actual administration of the diagnostic examination of the ESL composition courses. The tradeoff for attempting to conduct a study under normal circumstances in order to minimize the artificiality of highly controlled experimental conditions is that ideal experimental control suffers.
A second limitation to this study stems from the fact that subjects were informed prior to the administration of the second writing exercise that their responses were to be used for research purposes. The subjects' knowledge of this fact may have affected their performance on the second writing exercise.

The fact that there were fewer subjects at the advanced level than at the other two course levels may have resulted in the failure of the Scheffé test to detect significant differences between the advanced and intermediate groups on the basis of their holistic scores, T-unit indices, and error-free T-unit indices. Fewer subjects were available at the advanced level because most of the advanced subjects in the ESL courses did not meet the researcher's selection criteria that the subjects be new to the United States and to OSU. It may always be problematic to find large numbers of students at different levels who are homogeneous in terms of the criteria established by the researcher.

Despite the limitations that will occur in any empirical investigation, research of the kind reported in the present study is invaluable. The limited number of studies on topic variables have shown that aspects of topic variables influence writing performance in critical ways. If accurate writing assessments are to be made, many additional studies on topic effects are clearly in order.
APPENDIX A

TOPIC SHEET FOR DIAGNOSTIC EXAMINATION

You have the entire class period to write a well-organized, well-developed essay on the topic given below. Take five minutes or so to plan your composition before you begin to write, and allow at least five minutes at the end for making corrections. Your diagnostic will be evaluated on the basis of grammatical accuracy, syntactic variety, appropriate use of vocabulary, academic organization, and good development of your ideas.

In a recent news magazine, a famous educator argued that progress makes us lazy. Do you agree or disagree with this point of view? Explain why you believe that progress does or does not cause people to become more lazy or passive. Support your answer with specific reasons and examples.

NOTES
APPENDIX B

TOPIC SHEET FOR EXPERIMENTAL ESSAY

You have the entire class period to write a well-organized, well-developed essay on the topic given below. Take five minutes or so to plan your composition before you begin to write, and allow at least five minutes at the end for making corrections. Your diagnostic will be evaluated on the basis of grammatical accuracy, syntactic variety, appropriate use of vocabulary, academic organization, and good development of your ideas.

Every field of study has controversial issues. Debate over these issues often occurs among professionals in the field and leads them to conduct research in order to look for evidence to support one position on the issue over another or others. Choose a current controversial issue in your field of study. Discuss the controversy and explain your position on the issue, being sure to provide examples to support your position.

NOTES
APPENDIX C

STUDENT DATA QUESTIONNAIRE

Name:_________________________ Course:______ Qtr.____________

University ID:_______________ Instructor:_____________________

Check the appropriate info: 

Graduate-- Master Ph.D. College:__________________________

Undergraduate-- 1 2 3 4 Department:_______________________

Non-degree (special)______ Major:__________________________

Other - explain___________ Advisor:________________________

Native country/language:____________________________________

Columbus address:____________________ Phone:__________________

How long have you been in the United States?____________________

How long have you been at OSU?______________________________

Are any members of your family here in Columbus? yes no

If so, which members?______________________________________

Have you taken English 106, 107, 107-G, 108.01 or 108.02 at OSU? If
so, list the course(s), the quarter taken and the name of the

   teacher:_________________________________________________

Have you ever taken a writing course other than at OSU? If so, when
and where?______________________________________________

146
Please list the places and dates of any schools attended in the United States.

<table>
<thead>
<tr>
<th>School</th>
<th>Dates Attended</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Are you a Teaching or Research Associate this quarter? If so, list your working hours and briefly describe your duties.

Will you have any other kind of employment than the above this quarter? If so, list working hours and briefly describe duties.

What kinds of things will you have to write in your career as a student as a student or professional?

Is there any other information about you, either academic or personal that would be useful for your teacher to know?
APPENDIX D

INSTRUCTIONS FOR ADMINISTRATION
OF DIAGNOSTIC EXAMINATION

The Day Before

-- Announce that on Friday, Sept. 25, students will have the
diagnostic exam, which provides them with the opportunity to move
up to a higher level or to qualify (108.02).

-- Encourage all to attend. Announce that those who do not attend
will be marked as unexcused.

-- Put the following on the board and go over each point.
   - Bring dictionaries and pen or pencil.
   - Paper is provided.
   - Bring a picture I.D. (university I.D. or passport).
   - Arrive before the bell rings so that you can
     begin writing promptly.
   - You will have only 48 minutes to write.

-- Explain that they will not see the topic until the time of the
exam.

-- Please provide NO further instructions (i.e., do not remark on
organizational or grammatical points that may aid students).
During this time please make note of any questions that arise and your responses to them. Please make sure that I get this information.

**The Day of the Exam (Friday, Sept. 25)**

**GO OVER THIS INFORMATION THURSDAY AS WELL**

Please arrive to Denney with ample time before your class begins so that you can review the instructions and see me if you have any questions.

In your mailbox you will find an envelope containing:

- instructions
- a sheet for your remarks
- sealed topic sheets and paper packets with corresponding numbers
- extra paper and paper clips

Arrive to class a few minutes early and pass out exams only to students who either are listed on your roster or have written permission to take the diagnostic in your room. Instruct students to fill in the information on the folded sheet only (they must both print and sign their names). They should not write their names on the other pages. Tell them not to look at the question until instructed to do so (the topic sheet will be sealed). Please put the following information on the board:

- Do NOT open the exam.
- Fill in only the information on the folded sheet (print and sign your name).
- Do NOT write your name on the other pages.
- In front of you, you should have only your dictionary, exam, pen or pencil, and picture I.D.

-- Announce that extra paper is available if needed. Please be stingy with paper. Give out one sheet at a time and only when you see that it is needed.

-- Announce that no one is to leave early. If someone finishes ahead of time, s/he should read over the exam and wait until it is time for everyone to stop.

-- Announce that everyone must stop writing when instructed to do so. Explain that if they do not stop, their papers will NOT be considered for move-up.

-- Advise students to take five minutes to organize their thoughts before writing (there will be room on the topic sheets for them to make notes) and that you will announce ten minutes before the bell rings that ten minutes remain and that they should finish writing and leave time to revise.

-- When the bell rings, have students open the topic sheet. Read the question aloud and have students begin. If for some reason you are not ready to start at the sound of the bell, take notice of the actual starting time. Later, compensate with the same number of minutes after the bell rings at the end of the class. If this happens, please make note of it on the sheet provided for your remarks.

-- On the sheet provided for your remarks, jot down the names and approximate times of late arrivals.
-- As students write, go around to check I.D.'s, the name on their topic sheets, and your roster. If the student is not on your roster, make sure s/he has written permission to take the diagnostic in your class.

-- Carefully monitor students throughout the exam to ensure that no one refers to notes or talks to others, etc.

-- Ten minutes before the bell rings, announce that ten minutes remain and that students should conclude their papers and read them over for corrections.

-- At the sound of the bell, instruct students to stop writing. (BE FIRM: do not allow anyone to continue writing, despite his or her arrival time.) Have students fold the topic sheet around the exam and any extra sheets of paper they have used. Collect the exams and all other sheets of paper and return them to the envelope (paper clips are available to attach extra sheets). MAKE SURE THAT NO COPY OF THE TOPIC LEAVES THE ROOM!

-- Tell students that results will be available on Monday, Sept. 28, and that they must come to class to get the results. No results will be given out by phone.

-- If you cannot go immediately to your office to begin reading the exams, please leave your exams with me in 559. I will be available to help in reading exams and making referrals for move-ups. It is imperative that the exams be kept secure.

-- As you read the exams to make referrals, please do not make any marks on the exams. Make all comments on the referral sheets we have available.
Before 5:00 p.m. on Friday all exams should have been read and referrals made. Please, DO NOT take exams home with you. You will receive copies of the exams and the results Monday morning. If a student wishes to go over the exam with you, that’s fine. Just let me know the names of students who decide to do so.

Please make note of any situation that occurs which does not accord with the instructions outlined above.

It’s essential that these instructions be followed carefully. Please see me if you have any questions (559 Denney, 292-7281; home 262-1870). Thank you very much for your cooperation. It is greatly appreciated.
APPENDIX E

SHEET PROVIDED FOR INSTRUCTORS’ COMMENTS

Instructor____________________ This sheet is for your remarks
(late arrivals, out-of-the-
ordinary events, etc.).
APPENDIX F

INFORMATION REQUESTED ON OUTSIDE OF SEALED EXAMINATION

Name__________________________________________

                         Last                First

Signature________________________________________

Circle one:

Undergraduate          Graduate

Course________________________

Instructor_______________

Time____________________
APPENDIX G
MEMORANDUM TO STUDENTS

Subject: Writing exercise for Friday, October 16.
Date: October 14, 1987
To: 106, 107-G, and 108.02 students
From: ESL Programs

On Friday, October 16, you will be asked to write another in-class essay, similar to the one you wrote for the diagnostic exam at the beginning of the quarter. In the past, students have argued that the diagnostic topics they receive are too abstract and, therefore, do not allow them to demonstrate their actual writing ability. For this reason, we want you to write on a different kind of topic so that we can compare the writing samples and determine which kind of topic elicits the better writing performance. Although this writing exercise will be conducted exactly like the diagnostic exam, it will not have any effect on your present status in the ESL program. In other words, the results of this writing exercise are for research purposes only; they will not influence in any way your current course level or your class grade. This exercise is being considered as a regular classroom activity and, therefore, your attendance is expected as usual. Absences will be marked as unexcused. Please understand that your
identity will be kept anonymous in this study. After you write your response to this second question, the number on your paper will be matched with the number on your diagnostic exam and your name will no longer be kept on record. Only your samples of writing are of interest to the researcher in this study. If for some reason, however, you prefer not to have your writing samples included in the study, please inform your instructor. A decision not to have your writing samples included will be kept confidential and will in no way influence your grade in this course.

On Friday you will need to bring dictionaries and a pen or pencil (paper is provided). Be sure to arrive before the bell rings so that you can begin promptly. You will have only 48 minutes to write.

Your participation in this study will be greatly appreciated. It is our hope that the results will be of value to this program. Thank you for your cooperation.
APPENDIX H

INSTRUCTIONS FOR ADMINISTRATION
OF EXPERIMENTAL ESSAY

The Day Before

-- Announce that on Friday, Oct. 16, students will have the writing exercise that is for research purposes only.

-- Encourage all to attend. Announce that those who do not attend will be marked as unexcused.

-- Put the following on the board and go over each point.
  - Bring dictionaries and pen or pencil.
  - Paper is provided.
  - Arrive before the bell rings so that you can begin promptly.
  - You will have only 48 minutes to write.

-- Explain that students will not see the topic until the time of the exam.

-- Please provide NO further instructions (i.e., do not remark on organizational or grammatical points that may aid students).

-- During this time please make note of any questions that arise and your responses to them. Please make sure that I get this information.
The Day of the Exercise (Friday, Oct. 16)  GO OVER THIS

INFORMATION THE DAY BEFORE ALSO

-- Please arrive to Denney with ample time before your class begins so that you can review the instructions and see me if you have any questions.

-- In your mailbox you will find an envelope containing:
   - instructions
   - a sheet for your remarks
   - sealed topic sheets and paper packets with corresponding numbers
   - extra paper and paper clips.

-- Arrive to class a few minutes early and pass out exams. Instruct students to fill in the information on the folded sheet only (they must both print and sign their names). They should not write their names on the other pages. Tell them not to look at the question until instructed to do so (the topic sheet will be sealed). Please put the following information on the board.
   - Do NOT open the exam.
   - Fill in only the information on the folded sheet (print and sign your name).
   - Do NOT write your name on the other pages.
   - In front of you, you should have only your dictionary, exam, pen or pencil.

-- Announce that extra paper is available if needed. Please be stingy with paper. Give out one sheet at a time and only when you see that it is needed.
-- Announce that no one is to leave early. If someone finishes ahead of time, s/he should read over the exam and wait until it is time for everyone to stop.

-- Announce that everyone must stop writing when instructed to do so. Explain that if they do not stop, their papers will not be included in the study.

-- Advise students to take five minutes to organize their thoughts before writing (there will be room on the topic sheets for them to make notes) and that you will announce ten minutes before the bell rings that ten minutes remain and that they should finish writing and leave time to revise.

-- When the bell rings, have students open the topic sheet. Read the question aloud and have students begin. When you read the question aloud on Friday, you may see some confused faces. Just encourage students to make a decision about the topic and write. If they want to discuss a topic dealing with their undergraduate studies (if different from their graduate major), that’s fine. Please make note of the names of students who seem to take a long time to start to write. If for some reason you are not ready to start at the sound of the bell, take notice of the actual starting time. Later, compensate with the same number of minutes after the bell rings at the end of class. If this happens, please make note of it on the sheet provided for your remarks.

-- On the sheet provided for your remarks, jot down the names and approximate times of late arrivals.
-- Carefully monitor students throughout the exam to ensure that no one refers to notes or talks to others, etc.

-- Ten minutes before the bell rings, announce that ten minutes remain and that students should conclude their papers and read them over for corrections.

-- At the sound of the bell, instruct students to stop writing. (BE FIRM: do not allow anyone to continue writing, despite his or her arrival time.) Have your students fold the topic sheet around the exam and any extra sheets of paper and return them to the envelope (paper clips are available to attach extra sheets). MAKE SURE THAT NO COPY OF THE TOPIC LEAVES THE ROOM!

-- Please return the envelopes with all the exams to me after your class. Please DO NOT take exams home with you. I'll return to you the exams of those students who will not be included in the study.

-- Please make note of any situation that occurs which does not accord with the instructions outlined above.

It's essential that these instructions be followed carefully. Please see me if you have any questions (559 Denney, 292-7281; home 262-1870). Thank you very much for your cooperation. It is greatly appreciated
APPENDIX I
CALIBRATION SAMPLES

LEVEL SIX SAMPLES (**Note: Errors have been left uncorrected.)

Essay #258:

Seismic safety factor is one of the most important considerations in building design, specially in areas with high seismicity such as California, Mexico and Japan. Seismic forces are found by simulating the physical conditions that help or risk the security of a building when a seism occurs. Some of the most important conditions are soil and elasticity of the structure. Due to the uncertainty and lack of accurate experimental methods to determine these these physical conditions or to predict how intense the next seism is going to be, it is necessary to increase the seismic forces found by the simulation. The question is how high should this factor be.

Some engineers say safety factors are causing too much increase in construction costs but the performance of the structure does not increase accordingly, so they think safety factors can and should be reduced. On the other hand, when an earthquake occurs in a place where seismic safety factors are not conservative, there is a lot of destruction and lose of lifes, and other structures with considered reasonable safety factors collapse due to unknown factors when the building was constructed.
Essay #258 (cont.):

I think that while the real behavior and the real conditions of a structure exposed to seismic forces are not accurately known these safety factors must be high to compensate the ignorance. And the higher the ignorance of the site conditions, the higher the safety factor should be. If a person in the army die, his/her family is compensated with a considerable amount of money, so in buildings that can kill many people if it collapses, money must be invested to protect them. I do not believe that lifes can be changed for money, so higher costs are not enough excuse to reduce the safety factors. Later, when the unceirtanty could be lower, the safety factors could be reduced, but not today.

Essay #411:

In my opinion the role of progress, understood as the increase of productivity by using hight technology, has been to give more oportunities of leisure time for mankind. But it doesn’t mean that progress causes people to become more lazy or passive, because leisure time not necessarily implies lazy people; it depends on what people do in the leisure time.

Let me develop more my argument, I think that progress has been associated with a process of technological development that has produced a saving hand-work fenomenon in industry. It means that now people can produce more goods than in the past in the same period of time; or that people can produce now the same quantity of goods than in the past in a less period of time. Also along the history mankind has found other activities like art, recreation, literature, etc. that are
Essay #411 (cont.):

important for the full realization of men and women. But some centuries ago it was difficult for people to enjoy those activities because they needed to work hard to get basic products like food, clothes, housing, so forth. Progress has played an important role because people can get now more or the same quantity of basic products by working less time and it also implies that people has more free time to enjoy important activities like art and recreation, sports, etc. It is in this sense that progress has not cause people to become more lazy; by the contrary, progress implies more time to enjoy new opportunities. However the problem has been that persons have a high propensity to practice passive activities, these are the persons that we consider lazy, but the most of the people has found new activities in their leisure time and we have seen and increase in the performance of activities like dance, ballet, music, theatre, sports in general, travels and also education.

So progress can be associated with an increase in the leisure time and in the mankind opportunities, but not necessarily with more lazy or passive people.

Essay #179:

The current mainstream of public administration believes that public management can become a profession such as law, medicine and accounting. In reality, public administration can not become a profession.

According to the sociologists who have studied the professions, a profession is defined as an occupation having such characteristics as
Essay #179 (cont.):

unique techniques and knowledge, long-term training, professional organization and ethical standards.

Since 1930’s, many scholars have tried to establish the science of public administration with no success. Curricula at the leading school of public administration are diverse and there is no consensus on core courses of the field. The stock and depth of systematic knowledge is slim.

The training in school of public administration is not a requisite for the practice of public administration. According to some survey, only 3 percents of federal executive officials at the level of GS 16-18 have backgrounds on the training in public administration.

There is no professional organization representing the practitioners because the number of public administrators and the variety of their function are so huge as to prevent a organization from emerging. There are big gap between mainstream of scholars and practitioners. Public administration is evaluated in the context of hierarchy. A professional job is evaluated by peer group and ethical standard. Public administration obtains no legal recognition like medicine or law.

In sum, public administration is an occupation but cannot be a true profession. But nowadays the higher level position in federal government are occupied by professionals rather than those trained in public administration programs. The movement of professionals into executive position has made the practice of public administration
Essay #179 (cont.):

professional. They broaden their previous narrower perspective as they face complex political environment. They become 'administrator-professionals.'

The present situation that many professionals work in higher government positions and combine the professional technique and public administration, should be recognized by the school of public administration.

One response to current situation is to continue the diverse curricula emphasis and to bring up generalist public administrators, who are needed to integrate the highly specialized jobs in government. Another response is to increase the interaction with the practitioners and other professional schools. The former is continuing education for those who are or will be 'administrator-professional'. Meeting or conference with practitioners and internship in government will be helpful. The latter is to offer dual or joint degree in cooperation with other professional schools. In this case, public administration might be an education program to complement the professional educations, but it solve the frustration in locus or focus of public administration studies.

Essay #021:

Let's start with the definition of the word progress. Progress means improvement and because men are involved with progress, improvement focus to the way of human life. Progress has occurred in many parts of our every day-life and as far as I am concerned, I believe that (progress) provided us with both "Good" and "Bad" things.
Essay #021 (cont.):

I would like to write first about the "Bad" things. Well, I could hardly disagree with the famous educator that progress makes us lazy. If we examine our day-life now, I mean these days, and compare it with the way our parents lived we can easily see the difference. I would like to start my argument with the most basic thing man has to provide himself with, in order to be kept alive: food. What about it? Let's think about our kitchens and refrigerator. We hardly use nowadays our kitchen utilities. We have our "GOD" named microwave and his children named super-markets. We go once a week in one of them, we buy our food: frozen chicken, frozen cakes etc. and that's all. If, during the week run out of food, no problem, we can just order a piza or buy something from these FAST-FOODS. So eventually we don't have to bother with cooking, washing up etc. Of course I needn't write about the day of childhood, when the kitchen smelt so nice and mother spent the hole day preparing breakfast, lunch, dinner.

Another basic need we have, is to be clean. Our grandmother or even our mothers washed in a trade. What about today? Today we have the LAUNDRIES. We go there we put some coins and all of a sudden our work is finished, and we needn't spend our free time with such a boring activity.

But then how do we spend our free time? I should first of all ask if we have free time. I mean that this progress didn't appear all of sudden. Many people, trying to improve their way of living and serve themselves, got involved with this endless "project" of progress. They
Essay #021 (cont.):

worked and are still working hard, they even spend days and nights for improving progress. Those people I could certainly not accuse of lazyness.

Even those, including myself, could accuse of being lazy because of progress, because progress is very expensive and people have to work hard in order to get the money to buy the progressive things. By the end of the day they are so tired that they have no energy for any other activity than getting back home, take a bath, prepare something in the microwave and sit "lazy" in front of T.V. (the other great step of progress) and try to relax. They don't have the courage to go for a walk, visit a friend, they have telephone instead of it, or even go to a cinema or a theater. In this point of view progress made us not only lazy and passive but also unsocialable.

Well I could spend a hole day writing about progress but I don't have the time. The last thing I would like to mention is the progress made in Personal Computers and computers in general. Nowadays all our needs can easily be fulfilled by a microcomputer. We can pay our bills, we can entertain ourselves by playing with OUR NEW, PERFECT, FRIEND, we can type our work much more easy we can even "talk" with our friend with our computers. Of course progress has improved our lives in many other things health, education, libraries etc. but I generally agry with the educator and I disagree in the same time. I disagree because I wouldn't accuse progress of making us lazy and passive, I would better accuse man who made this progress in order...
LEVEL FIVE SAMPLES

Essay #403:

Undoubtedly during the twentieth century, a lot of new discoveries have provided us with more comforts and, certainly, have increased our standard of living. Many of these discoveries have substituted for the hand, and the feet of the man. For example, the washing-machine the dry-machine, the car, the computer etc. For example prior to the discovery of the washing-machine the women had to wash their clothes by using their hands. In addition the discovery of computer has provoked a real revolution.

Many people claim that all these discoveries have made the human race lazy. Well, in this statement there are two points of view. Undoubtedly nobody can deny that the effort, the men have to make in order to produce some good, has reduced. There are machines that are able to plow many acres in as much as the one percent of the time an ox and a farmer could plow the same land. Using a car or a plane one can travel thousands of miles in a few hours. Using a computer in accounting, for example, a lot of work can be done in as much as 5 percent of the time an accountant should need. But does this reducement of the effort mean that we have been lazy?

On the other hand all these machines cannot operate if the human mind doesn’t work. The farmer has to know how to use and how to repair the plow-machine, and to obtain this knowledge he has to use his mind. He has to learn the operation of the machine. Besides that a thousand of other people has to work both mentally and with their hands in order to produce this plow-machine. Some other people has to think how they
Essay #403 (cont.):

can improve the performance of that machine. The same holds for the cars, the planes and in most of all in the computers. In the computers a lot of people has to think how they'll program the computers.

Consequently the most men have to work harder, but using their mind in order to be able to obtain the benefits of the technology. The faster the progress of the technology moves on the more the man has to think.

Essay #80:

Failure Theories in Composite Materials

Owing to its development and its ever growing use in aerospace structure, composite material is being nowadays more thoroughly studied. A current controversial issue is the choice of a suitable failure criterion. A study conducted by AIAA composite material subcommite has shown that 90% of the designers uses maximum stress, maximum strain, and quadratic polynomial criteria. However these methods give, for the same specimen, different strength prediction; and researchers are nowadays conducting more research, experimental and analytical, in order to find a reliable failure criterion.

Composite material laminate is composed of lamina each in in turn composed of fibers embedded in the matrix material. Lamina may have different fiber orientations. Therefore, due to its complex composition, strength prediction of the laminate is also a complex problem. Some researchers use the maximum stress criterion which states that failure occurs when the stress, in anyone direction, reaches its maximum. Another criterion is the maximum strain criterion
Essay #80 (cont.):

which similarly states that failure occurs when the strain, in any one
direction, reaches its maximum. Other researchers believe that an
interactive failure criterion should be used to describe failure of the
composite material. The simplest interactive criterion would be the
quadratic criterion: $F_i^T + F_j^T + F_{ij} = 0$, where the factors $F_i$, $F_j$
are to be experimentally determined. However, the determination of the
factor $F_{xy}$ can not easily be determined experimentally and has to be
assumed. Accordingly different quadratic criterion has been used such
as Tsai-Wu, Hoffman, Chamis, Tsai-Hill.

A recent comparative study of these criteria has been conducted.
The comparative study shows that a difference up to 50 or 70% might
occurs between different strength predictions based on these different
failure criteria. In my opinion, the maximum stress, or maximum strain
criterion oversimplify the problem and a quadratic failure criterion
must be used. Further the Tsai-Wu quadratic criterion has been proved
by Dr. Tsai to be reliable in most designs. The Chamis criterion
assume different values of $F_{xy}$ in different quadrants which in my
opinion has no theoretical background and is used to better fit the
experimental data.

The issue is by no means closed and more studies must be conducted
in the future.

Essay #099:

The subject of Linear programming is very well known in
Optimization field, and the well known Simplex method is being used to
solve problems in this category. Many approaches are available to
Essay #099 (cont.):
simplify the calculations in simplex method. The most popular one is
the tablue procedure, where you put the problem in matrix form and try
to solve it step by step in a neat tablue. Another approach to
simplify the simplex method is called "Finite Improvement Algorithm".
This algorithm is, in fact, the simplex method in a new face. In
teaching the simplex method the tablue form was used for longtime as it
is easy to calculate in a tablue and the fact that in a tablue you can
follow all the calculation easily. On the other hand finite
improvement algorithm has less calculation to do as it calculate only
what is needed when it is needed, not the whole problem matrix as in
the tablue approach. One big disadvantage in the tablue approach is
that students will follow some rules in solving the problem and forget
the meaning of them as it become a routine calculation without feeling
what is really happening. This feeling can not be forgotten in finite
improvement alg. as it is the steps in solving the problem. It is
important to remember that the main issue here is to let students
understand what is happening in simplex method not only make some
routine calculation.

Essay #17 (This essay, although originally rated as a five, was
criticized by the raters and was, therefore, considered
a four during the calibration.)

Topic: Does progress make us lazy?

This is a controversial argument that has been discussed since the
rapid development came to the civilization. In my opinion, the more
Essay #17 (cont.):

progress we have, the more efficiency we obtain, that is progress will never make us lazy.

To clear this argument, we need examples which strongly support the aspect. As an illustration, the computer is the best choice. Today, people almost can not survive without the aid of computer, especially for those people working in the field of technology. With such a fantastic machine, every thing, therefore, becomes so easy and so solvable. However, it seems that the easiness can not satisfy the greediness of human being. It is the "easiness" which strongly stimulates the more progressive development for such a machine. This is, of course, from the psychological point of view. That is, the progress makes people ask such questions: why not better? Why cannot I make it perfect? Moreover, as the other people enjoy the result of progress, the people who make the progress will always say: How proud I am! I am the maker of the civilization! With such a psychology, no one can say the progress makes us lazy. As for the another explanation, let's ask a simple question, which is related to the history: If the progress makes us lazy, why does the human history develop in such a way that every thing has gained progress since the forming of the original human society? Probably no one will doubt that "laziness" will make retrogression; Thus, if progress makes people lazy, the modern civilization will never come to us, the human society will remain violent, bloody, hopeless! On the other hand, since the history has developed progressively, we can absolutely overthrow the argument in which people say that the progress make us lazy.
Essay #17 (cont.):

With no question, there are some people who become lazy since they gain the benefit from the progress. However, we should always remember that they are not the people who make the progress. But, they are just the few amount of people who do not possess the "normal psychology" for which they consider the "progress" is the best "stimulation" for themselves. Human development is clear. It verifies that "progress will make us progressive, not lazy", this truth will hold forever, this will also lead to the more progressive, civilized development of human society.

LEVEL FOUR SAMPLES

Essay #662:

It has been for a long time that people argue about the topic of where does the virus come from. After we have been struck by the herpes and the AIDS recently, this question is getting more attractive and more popular than before.

Generally, there are two different theories about the origination of the virus. The first one says that the virus is a certain kind of cell. People who agree with this theory believe that although the virus does not have some kinds of structures which exist in the normal cell, the virus is still a kind of cell in view of its capability of reproduction. The second theory says the virus is not a cell because it does not have some important parts of a normal cell. And the people who favor this theory explain that the virus is only a part of a cell and is not a complete cell.
Essay #662 (cont.):

Personally, I believe the virus a kind of a special cell. Although it lacks of some components when it is compared to a normal cell, the virus still can undergo most functions of a normal cell does. The reason of the virus being lacks of some parts of a normal cell is due to natural revolution and mutation. However, the virus is still a kind of cell and is not only a part of a cell.

Essay #08:

Experiencing the life in a lesser developed country, one can feel the non activeness and laziness that people usually live in a more developed country. I came to the United State of America nine months ago, from Lebanon, a country considered as having relatively not progressed in the last few years. Comparing my life style now to that I used to live, I can conclude that the existing facilities rendered me more lazy and passive. The effect of this progress is actually gradual and will become relevant at the long run when the person become attached to these facilities and feels that he can not manage without them.

The facilities that might make a person more lazy are quite a few, and can range from small things such as a rollant carpet in an airport, a television with remote control, or chairs with wheels in an office to major things such as the existence of fully automatic cars or a supermarket in which one can find almost all that he might needs. I remember my reaction when I first walked on a rollant carpet in the airport. "Oh! it is much easier". Since then I started to get more used to the facilities here. The television programs are well
developed and invite the person to seat for long hours on the T.V. set. Remember that the person doesn’t have to move to change channels, he can do this while sitting in his “Lazy Boy Chair”. Furthermore, sitting in my office on my chair to study, I rarely get up to find something in the room, since my chair can move on its wheels! I also don’t have to walk a lot when I go shopping the way I used to in my country. I only need to get into an automatic car, very easy to drive, with an air conditioner and windows that operate automatically. Once I get to the supermarket I can find everything I want from vegetables to grocery...

I believe that I become attached to these facilities and I found myself getting less active and more lazy. Imagine yourself, dear reader, if you can live without all these facilities that you become used and attached to, and tell me, didn’t they make more lazy?

Essay #604: (This essay, although originally rated as a four, was criticized by the raters and was, therefore, considered a five during the calibration.)

There are many methods used for the design of reinforced concrete structures. The most two famous of which are the Working Stress Design method (WSD) and the Limit Design method (LD). In the field of structural engineering, some engineers prefer the first method and others may prefer the second method for the design of concrete structures.

In order to make a decision to support one position over the another we should discuss the differences between these two methods.
Essay #604 (cont.):

In the first method (WSD), there is a basic assumption that the straining actions (internal forces) and the deformations (displacements, rotations) of the structure are linearly dependent on the applied load from zero load and up to the failure or break down of the structure, (see figure 1). This method neglects the non-linearity of the behaviour of reinforced concrete structures which occurred due to the following factors:

1. The non-linear response of concrete itself
2. The appearance of tensile cracks in the tension zone of the structure
3. The yielding of steel reinforcements at high load levels

![Graphs showing linear and non-linear load-deformation behavior](image)

**fig. 1** W.S.D.M.  
**fig. 2** L.D.M.

On the other hand the (LD) method takes into consideration the preceding factors. Of course it is more difficult and includes much mathematical effort than the (WSD) method, but it gives a more accurate
Essay #604 (cont.):

description of the behaviour of the structure (see fig. 2) and also
gives a more economic solution. Therefore, most of the modern codes of
practice for the design of reinforced concrete structures recommend to
use the Limit Design method although they admit using the Working
Stress Design method.

Essay #285:

Along with the industrialization and modernization, people could
enjoy the convenience that technology bring to them. However, such
progress doesn't mean that people could be more lazy than ever.
Conversely, people have to learn how to adjust themselves to this
technological society.

For example, the computerized libraries help those who need the
reference books to inquire data rapidly and efficiently. Thus, the
scholars and students don't have to waste time in inquiring but spend
more time in their research.

Moreover, machines really play an important role in the modern
life. It saves the time in housekeeping, transportation and
communication etc. Thus, we could spend our time and energe on more
advanced study and research. In the meantime, we still could enjoy our
leisure time.

In conclusion, progress doesn't cause people to become more laze.
On the contrary, it encourage people to more aggressive in order to
enjoy the delicate life.
LEVEL THREE SAMPLES

Essay #728:

There are two different issues about children's learning language in my field. One is that we do not have to teach children to learn language, another is that we must teach children to learn language.

In my opinion, I agree with the second issue that children should be taught to learn language. Since children do not understand the world very much, so we have to teach them to know the world step by step. And the best way to teach them is from learning language. For example, when we want to teach children to know "dog," we would say to him "This is a dog. This is a dog." repeatedly, and children would imitate us, then they know "what is dog." Children's language learning is a gradual move, and they need adults' teaching. It is our teaching that children began to know the world. Therefore I agree with the point of view that we must teach children to learn language.

In the field of Education, there are so many controversial issues about children's language learning. Of course, we could learn all of them, but I just could accept one point of view. It is my opinion.

Essay #155:

What's progress? It's ancestors' efforts which make lives much easier than the old days. Does progress make us lazy? Yes and No.

Yes. For most ordinary people, they become lazier and lazier. If they could control television channels at a distance, they will never go near to the TV set. If they could get upstairs on an elevator, they will never get there by stepping on the stairs even there is only one floor to go. People get lazier and lazier in physicals.
Essay #155 (cont.):

No. Because progress has contributed many great scientific techniques which have turned into many machines computers, etc. They can save us time. We don't need to spend a lot of time on routines. We don't need to spend a lot of time on our calculations for the data which were from many significant science experiments. So, we can work harder with our brains to design more precision instruments, to develop more wonderful thoughts. All these may contribute to an easy life for our offsprings. This is progress.

Progress make people lazy only in physical side, doesn't involve the efforts on our minds. So the progresses will go on and on and will never stop.

Essay #563:

In the field of administration, management of organization is very important. Administration cannot attain its purposes without efficient management of organization. But how to manage the organization is depends on how we understand the organization.

There have been two views about organization; organic view and mechanic view.

In the early times of the study, the point of view in which organization is considered as a machine had been supported by many author. Many people who have this viewpoint thought that it is desirable to make organization similar to the effective and efficient machine. For this reason they tried to throw away the emotional factor in organization scientific management is a good example for this point of view.
Essay #563 (cont.):

Considerably many authors are confident that organization is orgnic body. And when we study the organization, it is better to examine the emotional and personal aspects in the organization.

These two viewpoints are not reconciled yet. But I think that organic viewpoint is more powerful. If the purpose of the science is the understanding of reality, we cannot understand the real mean of organization. This is why I agree to the organic view.

But one benefit of mechanic view is that we can construct the ideal type such as bureaucracy. By the help of this point of view we can reorganize our real organization.

In any way, it is true that mechanic view is more prescrived and organic view about organization is more described. Two point of view have is very important implication to start organizational study.

Essay #011:

In my opinion, I agree with the educator that progress makes us lazy. Nowadays, the science technology makes a great progress that helps human beings to change and improve their lives. But, unfortunately, it also causes people to become more laze and passive. We can see the situation very clearly in daily life. There are four parts that I will discuss and analyse here; one is in food, the others are in clothing, house, and transportation.

The first, in food. We know that people rely on food to continue their lives. Men used to work very hard to make their own food. But, it has changed today. Some people just depends upon their heritage for food, because they have a great deal of money. They do not need to
Essay #011 (cont.):

worry about the weather, the seeds...and so on. Then, they can buy food with money. There is another case that scientists invent all kinds of machines for example: microwave, refrigerator...even some artificial food to minimize the time people spend on their food. That is why people become more lazy. Secondly, in clothing. Today, if men want to wear a suit, they just go to the shop and buy one. People lose their creativity and individuality. There will be a day that all human beings wear the same kind of clothes. That will be horrible. Thirdly, in house. House is a shelter which supply people a place to rest. Each person needs different kinds of houses. But, today, they all live in the apartments, the same structure and the same design. People do not even think that they should need more time to picture of what kinds of houses that they really want to live in. Finally, in transportation. In ancient time, when people had to move, they usually walked or rided a horse. But, nowadays, people won't get so many troubles. They only thing that they need to work out is to buy a car or a ticket. People have less exercise and become more passive than ever before. (Time is limited. I can develop too much!)

Recently, progress has caused people to fall backwards. They do not use their minds to creat, to invent and to think. They rely on what other people give them. It will be a great sorrow if people do not change this situation. All the human beings will live in a same lifestyle......
LEVEL TWO SAMPLES

Essay #152:

Mankind has made constant efforts to promote the welfare of him. As he has successfully tided over many difficulties before him until now, he has accomplished startling progress in the many areas such as technology, electronics, science, transportation system, communication system, etc.

Although I believe that nobody disagree with that modern science technology and sor forth have given human beings many benefits and have had us more convenient, I do not think that progress in the all areas has always done good for human beings.

I believe that progress has made people become more lazy. We can do many things in our works with help of modern machine such as computer and typewriter, electronic calculator, and so forth. We are used to convinient transportation system. We like better to ride cars than to work. Because of such thing, we are lazy to exercise. Therefore sometimes we meet plumy men and women in the street.

I have gone to the shopping center to purchase a T.V. set with my senior two weeks ago. He solicited me to buy one which has remote control system because it is more convenient than what has not remote control system. Of course it is convinient for me.

Essay #729:

Physics is has a long history compare with other sciences. It has been developed and changed as time went by. It’s a process to correct "mistakes" which used to be thought as a right answere and get a new
Essay #729 (cont.):

answer which has thought to be right by now. When an experimental result came out, but it didn’t agree with a theory or a concept which has been established, there always exists a debate among physicists.

For example, $\nu$ used to be thought as a particle without mass. This is a fundamental concept in particle physics. But, however, a group of European astrophysicists discovered that $\nu$ in the universe seems has mass as much as 20 eV. On this issue, physicists have had a debate for a few years. Some of them believe, the experimental result is reliable. We must make correction in particle physics. They thought it may cause a new revolution in physics. But others thought the experimental result is unbelievable. They take that the particle physics can explain other experimental result based on the concept that $\nu$’s mass is zero as their evidence. So they have been trying to find some mistakes in the experiment. By now, neither of them has found enough evidences to prove their theory is right. The debate is going on.

I’m a first year graduate student. I haven’t examined the details of either the theory or the experiment yet. So I have no right say my position on this issue. But it’s really a interesting problem. I hope I can join in the debate in the near future.

Essay #59:

The world has made a great progress in recent years. People can do almost every kind of daily things easily. It is obvious that the people can have a easy life if there are no strong motivations for them to be in a difficult circumstances. As a side-effect of the progress
Essay #59 (cont.):

of the society, many people will become more and more lazy, both mentally and physically.

The function of the brain will be developed by practicing. The more one uses his brain, the stronger its function will be. But, since we have more and more computer program in the daily life, we can do almost every things by the terminal of computer without thinking over it. I had the experience when I have the tutorial to the students in mathematical courses. Many students use the calculator as much as they can. As a result, they can hardly do very simple arithmatic. They have take a good advantage of caculator, but they lost the chances to practise to improve their mental function. For those students, they can do nothing about caculating without calculator. Gradually, they will be lazy with thinking.

The physical function of the human body also needs to practise. For example, if the people don't do much hard work, the function of their hands will be degenerated. In the modern life, people make most of the products by machine, automatically. They don't need to make products by their hands. Gradually, more and more people will be lazy to make something by their hands. Now, the physical exercises become more and more important since people lack of physical labour and their body become lazy.

Of course, the world will make progress forever.

However, as we pointed out above, it has a side-effect, that is making people lazy. The problem is that we should realize and recognize this and find some solutions to avoid this.
Essay #75:

Are Children Center in Classroom?

That children are center or teacher are center in classroom is a controversial issue in educational theory and practice. There are two positions on this issue. One is to take children as the important role in classroom. They suggest children learn mainly by doing. The source of knowledge comes from the interaction between children and environment. In the other word, it come from children's doing to object. Therefore, children play a significant role in the learning process. They learn something through touching object, pulling them, pushing them, dropping them and so on. They should raise questions, find the solutions, answer questions by themselves. Teacher in teaching process should be a guider and a helper. They are in charge of providing curriculum, equipment, facilities, and help children get themselves ideas. Another point of view is to take teacher as the main role in classroom. They think that teacher should be main source of knowledge. Children get knowledge from teacher. Only by teacher, the culture heritage can carry down generation by generation. Learning by children's doing would be waste much time. It is a way for human being's exploring the world but not for new generation to learn something from old generation. That is the two point of views on this issue.

My position on this issue is a middle-place. I agree with both of them. They should combine their points together, and creat a complete theory. For example, we must learn the cultural heritage by teacher. We can not repeat the way which our forefather have gone to explore
Essay #75 (cont.): everything by ourselves. We already know what kinds of clothes we should wear in cold winter and need not to try again. On the other side, children are dynamic organism. They should learn actively and vividly. They should master the method of how to explore new knowledge and world. Through their exploration, they can learn more and better than they are told by other people. Practice is the only resource of human being's knowledge. So action should be children's source of knowledge. Taking both of them. We should get a new theory.

LEVEL ONE SAMPLES

Essay #306:

Whether progress will make people lazy dicided by every person has a progressive mind or not. More science and technologies, more living convenience. But it doesn't mean more convenient that people should be more lazy to do something.

If one person who is combative and full of ideal, he would try his best to use the progress of his daily own to develop his ideal. Therefore, progress just support the necessary facilities for him and he works hard everyday. On the other hand, if everyone just makes his daydreams and enjoys that convenience, then more convenience makes more lazy people. They use machines to do everything they can and satisfy those achievements. They drive cars instead of walking. They use washing machines to wash dishes for them. That would be more lazy people produced by progress.
Consequently, the lazy owe to their lazy mind. Progress doesn't make people lazy.

Essay #546:

Audiology is a new profession field, it raised in last century. There are many controversial issues within the newly field. This field is related to human's health; every examination, every procedure or any theory which is applied to people must be carefully, so any tiny conflicts will be debated between audologists.

Although audiology has about one hundred years long in its history. Bone conduction testing still be discussed frequently. Because, there are many difference existing between manufacturers and clinicians. Like: difference procedures, difference calibration methods, and difference vibrators placements. So much difference result in unceasing debate from long time ago. It is said, dabating will lead to truth, it is correctly. But, long time's controversial is no good for the carrer of a field, it had better standize those methods and procedures. So audiology may serves people more.

"Where is the best placement" is one of most popular discussion tops. Many researches and experiments has done for this problems. In general, there have three place be placed vibrator most often, mastoid, forehead, and vertex. According to researches and experiments, vertex placement has the most advantages than others, so mastoid or forehead is considered now. In mastoid placement, there has some advantages, for example, higher sensitivity, less power needed and masking can be used; Although forehead placement has more realibility and stability
Essay #546 (cont.):

than mastoid placement, but it need more power than the former, and it
can not do masking when A-B gap is existing in the result.

How to resolve is important, I think that to set up the
standardization, is the only way to resolve this problem.
Essay #495 (This essay, although originally rated as a one, was
criticized by the raters and was, therefore,
considered a two during the calibration.)

I agree that progress does cause people to become more lazy.

Progress really changes our life. Due to progress, we don't have
to take heavy work anymore. We use the computer to work for us in
factory. We only hire few labors to control the factory and most of
the labors will lost their job. They don't need to work, because the
computer can work twenty-four hours a day. No one can do the work
better than computer. It never complain the low paid and long time
work. But where the labors go? They don’t work anymore and more
lazy. Everything is depended on the machine. They just push a botton
and the work will be done perfectly. Then everyone doesn't need to use
his brain to think and the brain will be useless. It is very terrible
and they enjoy their life and get more and more laze.

Above all, I agree that progress does cause people to become more
lazy and I really don’t want to see the lazy future.

Essay #556:

When we consider breathe, it seems to be sign of life. Breathe is
more important in dance because dance needs lots of energy and
well-developed inhale and exhale are required. In this matter
Essay #556 (cont.):

diaophragm plays an important role. Although many experts on dance agree with importance of diaphram, its function has been argued.

Some people merely say that when excercising oxygen become short and we need good breathe not appearing outside. Others say that the position of body greatly contribute to breathe. They insist that diaphram are attached to the abdomen, thorox area, spine, so that the position of body change the location of diaphram and the capability of breathe. If we stand, abdomen take tension and diaphram goes higher then we can more active breathe. From my experience I agree with latter opinion. When I am lying on side I feel press and there are vivid action of breathe owing to diaphram. When we are breathing, diaphram goes up and down and it change the space inside body. With such function of diaphram we can breathe whether successively or not.
### APPENDIX J

#### SAMPLE PAGE OF SCORING SHEET

<table>
<thead>
<tr>
<th>NAME</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>001</td>
<td>Y</td>
<td>N</td>
<td>076</td>
<td>Y</td>
<td>N</td>
<td>133</td>
<td>Y</td>
</tr>
<tr>
<td>002</td>
<td>Y</td>
<td>N</td>
<td>079</td>
<td>Y</td>
<td>N</td>
<td>137</td>
<td>Y</td>
</tr>
<tr>
<td>004</td>
<td>Y</td>
<td>N</td>
<td>081</td>
<td>Y</td>
<td>N</td>
<td>140</td>
<td>Y</td>
</tr>
<tr>
<td>005</td>
<td>Y</td>
<td>N</td>
<td>082</td>
<td>Y</td>
<td>N</td>
<td>141</td>
<td>Y</td>
</tr>
<tr>
<td>009</td>
<td>Y</td>
<td>N</td>
<td>083</td>
<td>Y</td>
<td>N</td>
<td>149</td>
<td>Y</td>
</tr>
<tr>
<td>012</td>
<td>Y</td>
<td>N</td>
<td>085</td>
<td>Y</td>
<td>N</td>
<td>163</td>
<td>Y</td>
</tr>
<tr>
<td>013</td>
<td>Y</td>
<td>N</td>
<td>086</td>
<td>Y</td>
<td>N</td>
<td>164</td>
<td>Y</td>
</tr>
<tr>
<td>015</td>
<td>Y</td>
<td>N</td>
<td>088</td>
<td>Y</td>
<td>N</td>
<td>169</td>
<td>Y</td>
</tr>
<tr>
<td>022</td>
<td>Y</td>
<td>N</td>
<td>103</td>
<td>Y</td>
<td>N</td>
<td>170</td>
<td>Y</td>
</tr>
<tr>
<td>024</td>
<td>Y</td>
<td>N</td>
<td>104</td>
<td>Y</td>
<td>N</td>
<td>176</td>
<td>Y</td>
</tr>
<tr>
<td>030</td>
<td>Y</td>
<td>N</td>
<td>105</td>
<td>Y</td>
<td>N</td>
<td>180</td>
<td>Y</td>
</tr>
<tr>
<td>031</td>
<td>Y</td>
<td>N</td>
<td>106</td>
<td>Y</td>
<td>N</td>
<td>183</td>
<td>Y</td>
</tr>
<tr>
<td>039</td>
<td>Y</td>
<td>N</td>
<td>108</td>
<td>Y</td>
<td>N</td>
<td>185</td>
<td>Y</td>
</tr>
<tr>
<td>051</td>
<td>Y</td>
<td>N</td>
<td>110</td>
<td>Y</td>
<td>N</td>
<td>220</td>
<td>Y</td>
</tr>
<tr>
<td>059</td>
<td>Y</td>
<td>N</td>
<td>111</td>
<td>Y</td>
<td>N</td>
<td>239</td>
<td>Y</td>
</tr>
<tr>
<td>063</td>
<td>Y</td>
<td>N</td>
<td>118</td>
<td>Y</td>
<td>N</td>
<td>259</td>
<td>Y</td>
</tr>
<tr>
<td>065</td>
<td>Y</td>
<td>N</td>
<td>123</td>
<td>Y</td>
<td>N</td>
<td>263</td>
<td>Y</td>
</tr>
<tr>
<td>066</td>
<td>Y</td>
<td>N</td>
<td>126</td>
<td>Y</td>
<td>N</td>
<td>268</td>
<td>Y</td>
</tr>
<tr>
<td>069</td>
<td>Y</td>
<td>N</td>
<td>129</td>
<td>Y</td>
<td>N</td>
<td>269</td>
<td>Y</td>
</tr>
</tbody>
</table>

190
## APPENDIX K

### MULTIPLE INDEPENDENT SCORING PLAN

<table>
<thead>
<tr>
<th>Folder Number</th>
<th>First Reader</th>
<th>Second Reader</th>
<th>*</th>
<th>Folder Number</th>
<th>First Reader</th>
<th>Second Reader</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>*</td>
<td>13</td>
<td>E</td>
<td>H</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>C</td>
<td>*</td>
<td>14</td>
<td>F</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>D</td>
<td>*</td>
<td>15</td>
<td>G</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>E</td>
<td>*</td>
<td>16</td>
<td>H</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>F</td>
<td>*</td>
<td>17</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>G</td>
<td>*</td>
<td>18</td>
<td>B</td>
<td>H</td>
</tr>
<tr>
<td>7</td>
<td>G</td>
<td>H</td>
<td>*</td>
<td>19</td>
<td>C</td>
<td>A</td>
</tr>
<tr>
<td>8</td>
<td>H</td>
<td>A</td>
<td>*</td>
<td>20</td>
<td>D</td>
<td>B</td>
</tr>
<tr>
<td>9</td>
<td>A</td>
<td>D</td>
<td>*</td>
<td>21</td>
<td>E</td>
<td>C</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>E</td>
<td>*</td>
<td>22</td>
<td>F</td>
<td>D</td>
</tr>
<tr>
<td>11</td>
<td>C</td>
<td>F</td>
<td>*</td>
<td>23</td>
<td>G</td>
<td>E</td>
</tr>
<tr>
<td>12</td>
<td>D</td>
<td>G</td>
<td>*</td>
<td>24</td>
<td>H</td>
<td>F</td>
</tr>
</tbody>
</table>
### APPENDIX L

#### CALIBRATION CHECK

<table>
<thead>
<tr>
<th>Exam Number</th>
<th>Calibration Score</th>
<th>1st Reading rater/score</th>
<th>2nd Reading rater/score</th>
<th>3rd Reading rater/score</th>
</tr>
</thead>
<tbody>
<tr>
<td>005</td>
<td>1</td>
<td>D / 1</td>
<td>G / 1</td>
<td></td>
</tr>
<tr>
<td>015</td>
<td>1</td>
<td>C / 1</td>
<td>F / 1</td>
<td></td>
</tr>
<tr>
<td>739</td>
<td>1</td>
<td>F / 1</td>
<td>A / 2</td>
<td></td>
</tr>
<tr>
<td>542</td>
<td>1</td>
<td>B / 2</td>
<td>E / 1</td>
<td></td>
</tr>
<tr>
<td>183</td>
<td>2</td>
<td>G / 3</td>
<td>B / 2</td>
<td></td>
</tr>
<tr>
<td>349</td>
<td>2</td>
<td>E / 2</td>
<td>H / 3</td>
<td></td>
</tr>
<tr>
<td>305</td>
<td>2</td>
<td>A / 2</td>
<td>D / 3</td>
<td></td>
</tr>
<tr>
<td>517</td>
<td>2</td>
<td>C / 4</td>
<td>F / 1</td>
<td>D / 3</td>
</tr>
<tr>
<td>503</td>
<td>2</td>
<td>D / 2</td>
<td>G / 1</td>
<td></td>
</tr>
<tr>
<td>323</td>
<td>3</td>
<td>F / 3</td>
<td>A / 2</td>
<td></td>
</tr>
<tr>
<td>081</td>
<td>3</td>
<td>B / 4</td>
<td>E / 2</td>
<td>C / 3</td>
</tr>
<tr>
<td>578</td>
<td>3</td>
<td>F / 2</td>
<td>D / 4</td>
<td>E / 4</td>
</tr>
<tr>
<td>361</td>
<td>4</td>
<td>D / 4</td>
<td>G / 3</td>
<td></td>
</tr>
<tr>
<td>606</td>
<td>4</td>
<td>E / 4</td>
<td>C / 6</td>
<td>G / 4</td>
</tr>
<tr>
<td>687</td>
<td>4</td>
<td>D / 3</td>
<td>B / 6</td>
<td>F / 5</td>
</tr>
<tr>
<td>557</td>
<td>5</td>
<td>G / 4</td>
<td>E / 4</td>
<td></td>
</tr>
<tr>
<td>269</td>
<td>5</td>
<td>A / 3</td>
<td>D / 4</td>
<td></td>
</tr>
<tr>
<td>065</td>
<td>5</td>
<td>B / 4</td>
<td>E / 4</td>
<td></td>
</tr>
<tr>
<td>414</td>
<td>6</td>
<td>H / 6</td>
<td>C / 4</td>
<td>Not read</td>
</tr>
<tr>
<td>039</td>
<td>6</td>
<td>G / 6</td>
<td>H / 6</td>
<td></td>
</tr>
</tbody>
</table>

---

192
APPENDIX M

SAMPLE ESSAYS USED FOR SCORING T-UNITS

Essay #414: 18 T-units; 8 error-free T-units

Progress produces many things which are very efficient in term of reducing time required in doing something. For example, the use of computer in writing reports, letters or in finding information from a data base will reduce time required significantly. Others high technology products will also help people to do something easier and faster and perfect.

The products of this progress are all around us, either we have it or we use it. The result is that we have much "spare" time. The question now is "What do we do with our spare time?"

Some people use their spare time doing nothing, some other people use it differently. For those who use their spare time doing nothing, we can say that the progress indeed makes them lazy. But I believe there are no many people in this category. Most people use their spare time by doing something they think is useful for them, like sports, hobbies, etc. The most significant advantage of spare time is by using it to be more productive. For example, a secretary normally types ten letters in one day. If the secretary uses a word processor, she (or he) can 'make' twenty letters in one day.
Essay #414 (cont.):

The progress does give people much more spare time, ease, convenience, which in turn will make people more productive./ It also gives people more time to think something that is useful for the human being and for the mankind./

Essay #539: 14 T-units; 2 error-free T-units

Thirty years ago, molecular biology started as a separate branch of biology./ This was a result of investigation of DNA structure./ The function of DNA is a carrier of inheritance./ Important question was how DNA could be responded for inheritance, and on which way it could determine protein synthesis./ Answer on this question was called "Dogma of molecular biology."

"Dogma of molecular biology" told that DNA molecule had to be transcribed to messenger RNA molecule which by itself had to be translated to a correct amino acid cue in protein./ It was told that it was the only way in which protein synthesis could occur, in every organism from virus to human body./ Some years ago, a special enzyme was found./ This enzyme, polymerase, can make DNA over mRNA as a template./ It was surprise to everybody and the end of "Dogma of molecular biology" in the same time./

Now, we know, that DNA and RNA synthesis can go in both ways, but only in some very simple organisms such as viruses are./ More convenient way is the first founded way/ but we never back to Dogma./ It is a good, because every dogma in the world is a bad thing./
Essay #005: 10 T-units; 1 error-free T-unit

Invention of high technology makes us feel very convenient in daily life. We can travel around the world in a week, not need 80 days, and we also can complete a lot of calculation in a few seconds by using computer, unnecessary to calculate from day to night and waste a lot of paper. But these conveniences will never make us lazy.

We should not forget the most important point that the creator of the progress is people. If people becomes lazy and stops the activities of creation, the progress of social civilization will also stop. People is the leader of progress, will never become passive in social development. The invention of car is a good example. Before people invented car, we have to walk for a long time from one place to another place, spend lot of time on travel, so reduce the time on activity of creation.

Essay #542: 18 T-units; 2 error-free T-units

My field of study is Educational Administration. This field provide persons to be administrators in schools. To be administrators, I must know many kind of things especially other problems issues and solved it. In my class, they are studying about the relation between courts and schools in term of the impact of law upon schools. My professor assigns students to find out these problems. I am not american students, so I know nothing about an idea for these issues. But I try to read in many books and then I found both relation between them. It is because courts try to control educational admission within the confine of the Constitutional Rights. But it has resulted to education leadership in case of lose
Essay #542 (cont.):

their power./ The best example that can explain this issue is the story of segregated in schools./ Many decades of year before, there was the problem about segregation to the black students in many schools in the United States./ This problem contrasts the democratic Constitution in term of the civil rights./ Black students had entered to study in a school as the same as white students./ But there were some segregation schools that were built,/ and it was illegal./ So this issue had to come to judicial by courts./
LIST OF REFERENCES


Brick, A. (1980). The CUNY Writing Assessment Test and the teaching of writing. Writing Program Administration, 4, 28-34.


Flower, L. S. (1988, May). Reading to write: The unseen work of task representation. Paper presented at The Ohio State University, Columbus, OH, Department of English.


Hudelson, S. (1984c, December). Developing writing abilities in a second language. Paper presented at the Ohio Conference of Language and Literacy for the Culturally Different Child, Columbus, OH.


