INFORMATION TO USERS

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

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

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

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame. If copyrighted materials were deleted you will find a target note listing the pages in the adjacent frame.

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

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.
A STUDY OF THE RELATIONSHIP BETWEEN SETS OF SECOND-LANGUAGE PROFICIENCY MEASURES AND READING COMPREHENSION MEASURES FOR ITALIAN TEXTS

The Ohio State University

Ph.D. 1982

Copyright 1982 by Scholl, Sheila Jolaine
All Rights Reserved
PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs or pages ______
2. Colored illustrations, paper or print ______
3. Photographs with dark background ______
4. Illustrations are poor copy ______
5. Pages with black marks, not original copy ______
6. Print shows through as there is text on both sides of page ______
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements ______
9. Tightly bound copy with print lost in spine ______
10. Computer printout pages with indistinct print ______
11. Page(s) ___________ lacking when material received, and not available from school or author.
12. Page(s) ___________ seem to be missing in numbering only as text follows.
13. Two pages numbered ___________. Text follows.
14. Curling and wrinkled pages ______
15. Other ___________________________
A STUDY OF THE RELATIONSHIP BETWEEN SETS OF SECOND-LANGUAGE PROFICIENCY MEASURES AND READING COMPREHENSION MEASURES FOR ITALIAN TEXTS

DISSERTATION

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

By
Sheila Jolaine Scholl, B.A., M.A.

The Ohio State University
1982

Reading Committee:
Professor Victor M. Rentel
Professor Melba D. Woodruff
Professor Neal F. Johnson

Approved By
Advisor
Department of Humanities
Education
Dedicated with love

to my parents,
A. Joseph and Maxine A. Scholl

and to my grandmother,
Mrs. James D. Berry
ACKNOWLEDGEMENTS

To Dr. Victor M. Rentel—for his willingness to chair this study, his professional insight, constructive comments, and encouragement—I am indebted and sincerely grateful. He imparts an enthusiasm and energy for research that motivated and sustained me throughout each stage of the study. The experience gained while working with Dr. Rentel has been enriching and memorable.

Professors Neal F. Johnson and Melba D. Woodruff, who served as members of my dissertation committee, shared views that contributed to the quality of my graduate studies in the areas of psycholinguistics and foreign-language education respectively. Their support is greatly appreciated.

Completion of the dissertation was in partial fulfillment of degree requirements; I wish also to thank Dr. Gilbert A. Jarvis, Dr. Edward D. Allen, and Dr. John J. Kennedy for guiding me toward this phase. I am particularly grateful to John Kennedy who stimulated my interest in research and injected valuable ideas into many discussions of design and analysis.

I wish also to express my appreciation to the students who participated in this study, to the instructors who granted their time and cooperation, and to those who were instrumental in arranging for data collection: Professors Luciano Farina and Albert N. Mancini of The Ohio State University, Professors Elaine McKee and Rufus Crane of the State
University of New York College at Buffalo, and Professor Peter N. Pedroni of Miami University. The assistance and thoughtfulness of Monika and Peter Pedroni are especially appreciated.

Many fond regards go also to close friends and colleagues. I thank Donatella Vigo-Glod who collaborated in the development of reading passages for use in the study, and Dr. Shirley J. Adams who assisted willingly in the scoring of recall summaries. Valued as well, is the genuine interest and friendship of Dr. Constance V. Hines, Bettye J. Myer, Teresa Karam, Marsha Mampieri, and Samuel R. Volpe.

Knowing the warmth and support of my family has been a constant source of security and strength throughout my life. To my mother and father, to my sisters—Malia and Renée, to my brothers—Michael and Rob, I extend my heartfelt thankfulness and deepest love. With affection also to John M. Konstantinidis for his understanding and belief in me.

ζων 
Γιάννη μου — το πιο μεγάλο εύχαστο.

iv
VITA

February 5, 1950........................ Born - Columbus, Ohio


1972............................... B. A., Miami University, Oxford, Ohio

1973............................... L'Università per Stranieri, Perugia, Italy (Summer enrollment)

1973 - 1977........................ Teaching Associate, Department of Romance Languages, The Ohio State University, Columbus, Ohio

1975............................... M. A., The Ohio State University, Columbus, Ohio

1977 - 1978........................ Research Associate, College of Humanities, The Ohio State University, Columbus, Ohio

1978 - 1979........................ Teaching Associate, Department of Romance Languages, The Ohio State University, Columbus, Ohio

1979 - 1981........................ Research Associate, College of Education, The Ohio State University, Columbus, Ohio

1981............................... Research Associate, Faculty of Humanities Education, The Ohio State University, Columbus, Ohio

PUBLICATIONS

FIELDS OF STUDY

Major Field: Foreign Language Education


Minor Field: Foundations and Research

  Studies in Research Design and Analysis. Professor John J. Kennedy
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEDICATION ................................................................. ii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS ................................................................ iii</td>
</tr>
<tr>
<td>VITA .................................................................................... v</td>
</tr>
<tr>
<td>LIST OF TABLES ....................................................................... x</td>
</tr>
<tr>
<td>LIST OF FIGURES ..................................................................... xii</td>
</tr>
</tbody>
</table>

## CHAPTER

### I. THE PROBLEM..................................................................... 1

Introduction ........................................................................... 1
Background .............................................................................. 3
Measures of Language Proficiency ............................................. 7
The Word-Boundary Task ......................................................... 10
Purpose of the Study ................................................................ 13
Definition of Terms .................................................................. 14
Significance of the Problem ....................................................... 15
Limitations of the Study ........................................................... 16

### II. REVIEW OF THE LITERATURE............................................. 17

Introduction ........................................................................... 17
Reading Comprehension and Language Proficiency .................... 17
  Current Views of Reading Comprehension ............................... 17
  Prior Knowledge as the Basis for Comprehension .................. 20
  Sources of Prior Knowledge and Reading Strategies ............... 25
  Word Knowledge, Syntactic Knowledge, and Inference .......... 36
Summary and Overview ............................................................ 45
Word-Boundary Task Experimentation ...................................... 46
  Methodological Aspects ........................................................ 46
  Measurement of Context Utilization ....................................... 57
Summary .................................................................................. 65
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>III. PROCEDURES</td>
<td>66</td>
</tr>
<tr>
<td>Population and Sample</td>
<td>66</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>68</td>
</tr>
<tr>
<td>Reading Comprehension Tests</td>
<td>68</td>
</tr>
<tr>
<td>Cloze Test</td>
<td>70</td>
</tr>
<tr>
<td>Tests of Language Proficiency</td>
<td>72</td>
</tr>
<tr>
<td>Textbooks</td>
<td>72</td>
</tr>
<tr>
<td>Syntax Test</td>
<td>75</td>
</tr>
<tr>
<td>Vocabulary Test</td>
<td>76</td>
</tr>
<tr>
<td>Word-Boundary Task</td>
<td>77</td>
</tr>
<tr>
<td>Materials</td>
<td>78</td>
</tr>
<tr>
<td>Procedures</td>
<td>80</td>
</tr>
<tr>
<td>Scoring</td>
<td>81</td>
</tr>
<tr>
<td>Data Collection Procedures</td>
<td>81</td>
</tr>
<tr>
<td>Main Study Reliability Assessments</td>
<td>84</td>
</tr>
<tr>
<td>Design and Statistical Analyses</td>
<td>85</td>
</tr>
<tr>
<td>IV. RESULTS AND DISCUSSION</td>
<td>89</td>
</tr>
<tr>
<td>Preliminary Analyses</td>
<td>89</td>
</tr>
<tr>
<td>Word-Boundary Task Experiment</td>
<td>90</td>
</tr>
<tr>
<td>University Effects</td>
<td>91</td>
</tr>
<tr>
<td>The Task Factor</td>
<td>96</td>
</tr>
<tr>
<td>Task by Task Sequence Interaction</td>
<td>99</td>
</tr>
<tr>
<td>Interaction of Task with Context Clueing and Boundary Markers</td>
<td>101</td>
</tr>
<tr>
<td>Context Utilization and the Word-Boundary Task</td>
<td>108</td>
</tr>
<tr>
<td>Relationships Between Measures of Language Proficiency and Reading Comprehension</td>
<td>112</td>
</tr>
<tr>
<td>Analysis and Interpretation of Between-Sets Correlations</td>
<td>112</td>
</tr>
<tr>
<td>Follow-up Procedures to the Canonical Correlations</td>
<td>116</td>
</tr>
<tr>
<td>Summary</td>
<td>123</td>
</tr>
</tbody>
</table>
## CHAPTER

<table>
<thead>
<tr>
<th>V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS</th>
<th>124</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview...........................................</td>
<td>124</td>
</tr>
<tr>
<td>Summary of Procedures............................</td>
<td>125</td>
</tr>
<tr>
<td>Summary of Findings................................</td>
<td>127</td>
</tr>
<tr>
<td>Review of Experimental Outcomes................</td>
<td>127</td>
</tr>
<tr>
<td>Review of Correlational Findings................</td>
<td>131</td>
</tr>
<tr>
<td>Recommendations for Further Research...........</td>
<td>133</td>
</tr>
</tbody>
</table>

BIBLIOGRAPHY............................................. 136

APPENDICES

| A. MATERIALS FOR MEASURING READING COMPREHENSION | 146 |
| B. TESTS OF SECOND-LANGUAGE PROFICIENCY........... | 156 |
| C. WORD-BOUNDARY TASK MATERIALS.................... | 164 |
| D. CHARACTERISTICS OF THE SAMPLE BY UNIVERSITY AND CLASS | 171 |
| E. DESCRIPTION OF CANONICAL ANALYSIS............... | 181 |
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.</td>
<td>Summary Description of Textbook Material</td>
<td>74</td>
</tr>
<tr>
<td>3.2.</td>
<td>Vocabulary Breakdown</td>
<td>75</td>
</tr>
<tr>
<td>3.3.</td>
<td>Syntax Test Reliabilities Based Upon Pilot Study Data</td>
<td>76</td>
</tr>
<tr>
<td>3.4.</td>
<td>Vocabulary Test Reliabilities Based Upon Pilot Study Data</td>
<td>77</td>
</tr>
<tr>
<td>3.5.</td>
<td>Description of Word-Boundary Instrument Design</td>
<td>79</td>
</tr>
<tr>
<td>4.1.</td>
<td>Summary of the Repeated Measures ANOVA by Context Clue, Boundary Markers, and Task Sequence</td>
<td>97</td>
</tr>
<tr>
<td>4.2.</td>
<td>Means and Standard Deviations for Word-Boundary Task Performance by Sequence and Task</td>
<td>98</td>
</tr>
<tr>
<td>4.3.</td>
<td>Descriptive Statistics of WBT Performance by Context Clue, Boundary Markers, and Task</td>
<td>103</td>
</tr>
<tr>
<td>4.4.</td>
<td>Means and Standard Deviations of Difference Scores by WBT Form</td>
<td>109</td>
</tr>
<tr>
<td>4.5.</td>
<td>Intercorrelations Among Measures of Language Proficiency, WBT Measures, and Measures of Reading Comprehension</td>
<td>114</td>
</tr>
<tr>
<td>4.6.</td>
<td>Summary of Canonical Analyses Between Language Proficiency Measures and Measures of Reading Comprehension</td>
<td>115</td>
</tr>
<tr>
<td>4.7.</td>
<td>Canonical Weights and Structure Coefficients of Left and Right Set Variables Associated with the Significant Pair of Canonical Variates (Roots) in Each Analysis</td>
<td>117</td>
</tr>
<tr>
<td>4.8.</td>
<td>Summary of Tests of Significance for Increments in Squared Multiple Correlations Between the Predictor Variables and Recall After Partialling Out Teacher-Class Effects</td>
<td>120</td>
</tr>
<tr>
<td>4.9.</td>
<td>Summary of Tests of Significance for Increments in Squared Multiple Correlations Between the Predictor Variables and Recognition After Partialling Out Teacher-Class Effects</td>
<td>122</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>D.1.</td>
<td>Descriptive Work-up of Personal Data for Total Sample and by University</td>
<td>173</td>
</tr>
<tr>
<td>D.2.</td>
<td>One-Factor MANOVAs on Seven Language Measures by University and by Class</td>
<td>174</td>
</tr>
<tr>
<td>D.3.</td>
<td>Univariate Analyses of Variance for Seven Language Measures by University and by Class</td>
<td>175</td>
</tr>
<tr>
<td>D.4.</td>
<td>Descriptive Statistics of Seven Verbal Measures by University</td>
<td>176</td>
</tr>
<tr>
<td>D.5.</td>
<td>Descriptive Statistics of Seven Verbal Measures by Class</td>
<td>177</td>
</tr>
<tr>
<td>D.6.</td>
<td>Summary of the ANOVA by University, Context Clueing, Boundary Markers, Task Sequence, and Task</td>
<td>178</td>
</tr>
<tr>
<td>D.7.</td>
<td>Descriptive Statistics of WBT Performance by University and Context Clue</td>
<td>180</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td>2.1.</td>
<td>Overlap of Linguistic Constraints</td>
<td>59</td>
</tr>
<tr>
<td>2.2.</td>
<td>WBT Performance Conditions - Random Version</td>
<td>61</td>
</tr>
<tr>
<td>2.3.</td>
<td>WBT Performance Conditions - Coherent Version</td>
<td>62</td>
</tr>
<tr>
<td>2.4.</td>
<td>Projected WBT Performance Continuum</td>
<td>63</td>
</tr>
<tr>
<td>3.1.</td>
<td>Three Between-One Within-Subjects Design Arrangement</td>
<td>87</td>
</tr>
<tr>
<td>4.1.</td>
<td>University by Context Clue Interaction</td>
<td>92</td>
</tr>
<tr>
<td>4.2.</td>
<td>Task by Task Sequence</td>
<td>100</td>
</tr>
<tr>
<td>4.3.</td>
<td>Task Sequence by Task</td>
<td>100</td>
</tr>
<tr>
<td>4.4.</td>
<td>Context Clue by Boundary Markers by Task Interaction</td>
<td>102</td>
</tr>
<tr>
<td>4.5.</td>
<td>Context Clue by Boundary Markers by Task: Another View</td>
<td>106</td>
</tr>
</tbody>
</table>
CHAPTER I
THE PROBLEM

Introduction

In an effort to understand the cognitive processes governing language acquisition and use, many psycholinguists have turned their attention to reading, a most complex form of information processing. Many current psycholinguistic theories concerning reading no longer view it as being a passive skill, but one that demands active, cognitive effort on the part of the reader. Models based upon such theories imply that reading is a problem-solving activity. These models not only assume the reader's knowledge of the language (syntax, semantics), but also portray the reader as being engaged in utilizing strategies such as anticipation, hypothesis-testing, and inference. The ability to manipulate and integrate these highly developed skills turns upon the simultaneous workings of elements inherent in language and interactions between reader and text.

Readers differ in their ability to interact with and construct meaning for text. Furthermore, reading comprehension entails component skills and processes, the importance and necessity of which may be perceived differently by different individuals. It is possible that skilled readers resort to lower levels of processing (feature, letter, word identification) only when hypotheses about higher-level orders (phrase,
clause, sentence information) have failed, while less proficient readers may operate at lower processing levels and find it difficult to transfer meaning to higher levels. But, there is also the likelihood that all readers make trade-offs among the various sources of information and that fluency is a matter of monitoring and synthesizing available cues effectively. Individual preferences for certain processing levels, or individual differences in the ability to synthesize information from these levels may be due, in part, to knowledge of the language, and in part, to the use of this knowledge.

While knowledge of the language is considered a prerequisite of reading, sheer knowledge of a language does not assure comprehension. Similarly, a reader may be very efficient in reading skill and strategies, but unable to employ them when, as will be shown below, prior knowledge of language or subject matter is lacking. although a balance of linguistic knowledge and use of prior knowledge may be desirable, an imbalance does not necessarily preclude comprehension. Proficiencies in one aspect may compensate for deficiencies in the other resulting in comparable comprehension outcomes.

The nature of the apparent relationship between language knowledge and its use remains elusive. Their respective dependencies, as well as their relative and cumulative contributions to the process of reading comprehension have not been thoroughly investigated.

The present study proceeded from two related objectives. The first was to explore the joint contributions of selected measures of second-language proficiency and reading strategies to the comprehensibility of texts written in Italian. Language proficiency was assessed in terms of
knowledge of vocabulary and syntactic rules. Reading strategies included the ability to infer meaning and context utilization—the use of language cues supplied in the form of contextual constraints. The inclusion of context utilization in the design of the study, however, was contingent upon the ability to measure that construct. Therefore, the second objective of this study was to investigate the construct validity of an instrument designed to measure context utilization.

It was feasible to pursue this line of inquiry with second-language learners for several reasons. First, the theories of reading comprehension relevant to this research have served as bases for several studies in the second-language domain. Secondly, the second-language classroom provides rather appealing conditions in which to examine individual differences in language proficiency because, generally, exposure to the target language is confined exclusively within the classroom setting and is limited to the content covered in the textbooks. In addition, research in second language demonstrates that some mature readers of native language can and do transfer their acquired skills and strategies to reading in a second language. Moreover, they may develop new strategies in second-language learning for coping with reading as a problem-solving activity.

Background

According to many current theories of reading comprehension, there is a definite relationship between the reader's prior "cognitive" knowledge and the ability to understand written text. Based on the assumption that individuals use prior knowledge when trying to make sense of new information, comprehension has been defined as "relating new
experience to the already known" (Smith, 1975; p. 10), and has been viewed as a "constructive" process involving "the ability to impose one's present knowledge on the text... in order to represent it and obtain information from it" (Glaser, 1979; p. 7).

The knowledge that the reader possesses has been subsumed by such rubrics as "cognitive structure" (Ausubel et al., 1978), "schemata", "frames", "scripts" (Bartlett, 1932; Minsky, 1975; Schank and Abelson, 1977), "nonvisual" or "nonsensory" information (Smith, 1975; Rumelhart, 1977) and prior contextual or background knowledge (Bransford and Johnson, 1972). Although some of these terms are more domain-specific than others, each makes reference in some way to information or knowledge that "is carried around by the reader all the time; it does not go away when the lights go out" (Smith, 1975; p. 5). This knowledge may be general or specific (Schank and Abelson, 1977). Thus, it may include knowledge of the world or environment, prior learnings or beliefs, memories for previous personal experiences, past events or situations, and so forth—in short, anything that represents "given" knowledge. For the obvious reason that the possibilities regarding an individual's knowledge are limitless, the extent to which background knowledge influences one's understanding of text cannot be determined. But, aspects of this knowledge can be isolated and examined in relation to reading comprehension.

A number of experimental studies conducted with native speakers have demonstrated the facilitative effects of generic knowledge (prior knowledge of context, relevant frameworks, and knowledge of everyday themes and events) on the comprehension of thematically obscured reading passages (Bransford and Johnson, 1972; Gordon, Hansen and Pearson, 1978;
Comparable results have been found in second-language settings whether the context of the passages was intentionally obscure (Adams, 1981) or straightforward (Omaggio, 1978; Labarca, 1979; Mueller, 1979). These studies suggest however, that proficiency in a target language still accounts for a significant amount of variance in comprehension, even when contextual cues are provided (Mueller, 1979), or appropriate scripts are activated (Adams, 1981). As Adams points out:

A second-language reader...possesses an imperfect knowledge of the language, so more information and cues are needed in order to predict [context] (p. 4).

Both Adams and Mueller concluded from their findings however, that if context can be generated easily from one's knowledge of linguistic cues, as might be the case for high proficiency groups, additional paralinguistic cues may be superfluous. Therefore, while relevant contextual knowledge may be a necessary component for meaningful reading, this knowledge can be supplied by other sources of knowledge within the reader's repertoire.

Many of these sources of knowledge are language specific, and may include information about the words and structure of a particular language as well as knowledge of its orthographic, syntactic, and semantic properties. In the case of a second language, acquired within a classroom setting, these knowledge sources are prior learnings that constitute only part of the "background knowledge" that the reader brings to a particular reading. It is an essential part, however, and one that may be of more practical importance to second-language readers than a broad contextual
framework for several reasons. One is that broad contextual knowledge can reduce only so much uncertainty within a text. Indeed, if the contextual knowledge is so complete that no further information can be gained, "reduction of uncertainty" or "comprehension" is simply not an issue (Garner, 1962; Smith, 1975). When reading for information, however, a broad contextual framework is not sufficient information to enable a reader who has no knowledge of Greek, for example, to understand text written in that language. On the other hand, in the absence of specific background knowledge, some degree of proficiency in the language may permit the reader not only to grasp the gist of the message, but to infer particulars within the general context as well. In addition, while readers may always have the necessary contextual knowledge at their disposal, often some stimulus signalling the reader to access this knowledge is required. Readers cannot be totally reliant on external stimuli of the type experimenters furnish, but they can depend upon internal sources of linguistic knowledge for triggering relevant scripts or schemata, presupposing some degree of proficiency in the language.

The phrase "some degree of proficiency" repeated in the preceding paragraph is necessarily vague. As a first step toward further specificity, the relationship between reading comprehension and language proficiency must be clarified. Strong bivariate correlations between various aspects of language proficiency (vocabulary or syntax measures, for example) and reading comprehension have been reported, however, multiple correlations have not been documented that take into consideration several sources of knowledge about language. Examining these relationships with isolated language proficiency factors not only distorts the view of
reading comprehension as a process involving the merging of many sources of knowledge available within the reader, but also tends to exaggerate the importance of individual proficiency factors. Since language proficiency implies knowledge of several characteristics of language, as well as skills in using that knowledge, correlational methods should be employed that reflect the multivariate nature of both language proficiency and reading comprehension. Such methods might provide a more conservative assessment of the contributions to reading comprehension that can be attributed to a reader's knowledge of language and strategies for using that knowledge.

Measures of Language Proficiency

In this study, the readers' second-language background was assessed through measures of language proficiency that included the knowledge of discrete language components and also the readers' ability to integrate those components. A rationale for the selection of specific proficiency measures is presented in this section.

Although there are no precise guidelines on how to measure language proficiency, precedents have been set. Traditional tests of second-language proficiency, still in use today, are divided into separate sections for testing discrete components such as phonology, vocabulary, and syntax, and skills (e.g., speaking, listening, reading, writing) of language. While subscores from each division can be used for diagnostic purposes, the total score ostensibly represents an individual's second-language proficiency. Thus, it is a score that reflects the compilation of distinct language skills and components. One of the merits of this method of testing is that judgements concerning an individual's command of a
language are not based solely upon performance in just one area of lan­
guage, nor upon a single test that may be biased or unreliable.

This "discrete-point" approach (Carroll, 1961) to measuring language
proficiency has been criticized, however, because as Oiler (1979) argues:

... crucial properties of language are lost when its
elements are separated. The fact is that in any system
where the parts interact to produce properties and
qualities that do not exist separately, the whole is
greater than the sum of its parts. If the parts cannot
just be shuffled together in any old order--if they must
rather be put together according to certain organiza­
tional constraints--those organizational constraints
themselves become crucial properties of the system which
simply cannot be found in the parts separately (p. 212).

As an alternative to test batteries aimed at composites of language
proficiency, Oiler suggests integrative or "pragmatic" language tests
that reflect the use of language under more natural circumstances. To
be included in this particular class of tests, Oiler specifies "two na­
turalness criteria" that a language task must meet:

... (1) It must require the processing of temporal se­
quences of elements in the language constrained by the
normal meaningful relationships of such elements in
discourse, and (2) it must require the performer of the
task to relate the sequences of elements to extralin­
guistic context via pragmatic mappings. More succinctly,
pragmatic tasks require time constrained processing of
the meanings coded in discourse (1979, p. 262).

One test that fulfills these requirements is the "cloze" procedure
that entails the restoration of missing elements in a message. Original­
ly a test of readability (Taylor, 1953), the utility of the cloze tech­
nique for measuring reading comprehension (Bormuth, 1967) and overall
language proficiency (see Oiler and Perkins, 1980) has also been acknow­
ledged. Moreover, cloze task performance lends empirical support for
the theory that readers employ strategies of anticipation, prediction, and inference when trying to solve problems of comprehension.

Although all of the precise skills measured by the cloze procedure have not been identified (Oiller and Conrad, 1971), performance appears to depend upon a sensitivity to the linguistic regularities and contextual information within a passage that create a "grammar of expectancy" (Oiller, 1979) for the respondent. Bialystok and Howard (1979) concluded from their research that "inferencing is an integral component in the solution of cloze tests" (p. 11). These researchers defined the "strategy of inferencing" as "the ability to maximally exploit available information sources to arrive at new insights into unknown aspects of the target language" (p. 3). Carroll (1972) questions the validity of the cloze technique as a single measure of reading comprehension and also maintains that cloze tests following an every nth-word deletion pattern do not "permit measuring the degree to which the individual comprehends particular lexical or grammatical cues, or possesses a knowledge of specified linguistic rules" (p. 19). But, he does agree that performance on cloze tests "probably depends to a considerable extent on inferential processes" (p. 19).

A cloze test was employed in this study as an indicator of subjects' "ability to infer meaning" from written text. This term was used primarily to distinguish the more "integrative" measure of language proficiency from two other measures that were obtained through discrete-point tests, namely, vocabulary knowledge and syntax knowledge. All three of the measures named have been implicated as important factors in reading comprehension, and are discussed further in Chapter II.
The Word-Boundary Task

A problem-solving task that bears a resemblance to the cloze procedure in terms of the strategies called upon during performance, is the word-boundary task (Klein and Klein, 1972). The task is designed so that words are printed with equal spacing between letters and across words. From this uniformly spaced array of letters, the respondent is required to identify word units. Many variations of the word-boundary task (WBT) have been employed experimentally primarily to examine the use of contextual information as a prerequisite for generating expectancies about text (Klein, Klein, and Vigoda, 1972; Klein and Klein, 1973; Klein, Klein, and Bertino, 1974; Klein, 1977; Schwartz, 1977). The procedures followed in these studies are essentially the same. In general, in a very brief time period, subjects are instructed to identify words within passages varying in terms of coherence or the presence of contextual constraints. Often subjects are presented with coherent passages and random versions of these passages, in which word order is randomized but orthographic constraints are preserved. Since the task also involves factors such as motor coordination and attention span, the random version serves as a baseline from which comparisons in performance can be made. Results have shown consistently that performance on the WBT is enhanced with increased constraints and greater approximations to coherent canonical text. Evidence from this research suggests that reliance on the structure imposed by orthographic, syntactic, and semantic constraints allows predictions to be made about future aspects of a message. According to Klein and Klein, (1973) the "word-boundary task thus--in the difference between the random passage and the passage with contextual
information—reflects the use of contextual information to maximize word predictions" (p. 400). An alternative explanation for the facilitative effects of context is provided by a schema theory of language comprehension. As Adams and Collins (1977) point out:

... a text only provides directions for the listener or reader as to how he should retrieve or construct the intended meaning from his own, previously acquired knowledge. The words of a text evoke in the reader associated concepts, their past interrelationships and their potential interrelationships. The organization of the text helps him to select among these conceptual complexes. The goal of schema theory is to specify the interface between the reader and the text—to specify how the reader's knowledge interacts and shapes the information on the page and to specify how that knowledge must be organized to support the interaction (p. 5).

Despite strong evidence that the increment in performance on coherent formats of the WBT indicates context utilization, only a few studies have suggested how a measure of context utilization might be obtained from observed performance on the WBT (Klein, Klein, and Bertino, 1974; Schwartz, 1977). More thorough investigation is needed to establish the construct validity of this instrument. Towards this end, the word-boundary task requires further clarification.

Two factors in particular that have been found to affect WBT performance are markers placed at phrase or sentence boundaries, and a contextual framework provided prior to the word identification task. In a series of experiments with the WBT, Klein, Klein, and Hildum (1974) found that the provision of additional information to the paradigm, such as phrase and sentence boundaries, "maximized" performance. They concluded that "language users may rely on whatever cues are available to provide predictability" (p. 340). Following their lead, Schwartz (1977)
investigated the effects of "thematic" clues on WBT performance in two experiments. In one instance he found that clues about the context of a passage made available prior to the word-identification task enhanced WBT performance. Based on the evidence provided by these researchers, sentential boundary markers and contextual clues were incorporated into the design of this study. In addition, as in all previous experiments conducted with the WBT, the sequence of presentation of the random and coherent passages was also controlled.

The work of Klein and Klein and their colleagues lends support to "active" and "interactive" models of reading in which predictive mechanisms are included (Neisser, 1967; Goodman, 1967; Smith, 1971; Gibson and Levin, 1975; Rumelhart, 1977). Their research also raises specific questions for further investigations such as "the conditions that determine whether subjects will utilize various types of contextual cues" and "the examination of individual differences in the use of contextual information" (Klein and Klein, 1973, p. 405-406). Theoretical principles and empirical evidence both suggest that measures of word-boundary task performance may provide insights into these two research problems. Valid measures of context utilization defined in terms of a word-boundary task, however, can be obtained only after further clarification of the methodology associated with that instrument.
Purpose of the Study

The purpose of the present study was threefold:

1. to examine the nature and strength of the relationship between measures of language proficiency, including (i) vocabulary knowledge, (ii) syntax knowledge, and (iii) the ability to infer meaning, and measures of reading comprehension of a second language;

2. to investigate the extent to which context utilization can be measured validly by means of a word-boundary task. This line of inquiry entailed (i) an experimental assessment of the methodological artifacts associated with the instrument, and (ii) an evaluation of the construct validity of performance on a word-boundary task as a measure of context utilization. Specifically, the following questions were addressed:

   - Is performance on a word-boundary task affected by the inclusion of such variables as (i) verbal contextual clueing, (ii) sentential boundary markers, and (iii) the presentation sequencing of random and coherent versions of a passage?

   - Are there differences in performance on a word-boundary task associated with differences in reading comprehension and second-language proficiency?

3. to explore the nature and strength of the relationship between measures of language proficiency, including (i) vocabulary knowledge, (ii) syntax knowledge, (iii) the ability to infer meaning, and (iv) context utilization and measures of reading comprehension of a second language.
Definition of Terms

Terms to be encountered in this study are defined as follows:

**Reading Comprehension** is defined by measures of recall and recognition obtained from tests administered after the reading of two passages written in the target language. **Recall** is a measure of the number of main ideas contained in summaries (written in English) of each reading passage. **Recognition** is measured by scores on two 5-item multiple-choice tests based on each reading passage.

**Language Proficiency** is defined in terms of knowledge of Italian syntax and vocabulary as measured by performance on tests developed by the researcher, and in terms of the ability to infer meaning and context utilization. The **Ability to Infer Meaning** is that ability defined by performance on a cloze passage prepared in the target language. **Context Utilization** is the ability to use the natural contextual constraints of written language. This strategy was defined operationally as performance on a word-boundary task adapted in the target language.

**Word-Boundary Task** is a device developed by Klein and Klein (1972) to assess a subject's use of syntactic and semantic information. The materials consist of passages printed with equal spacing between letters and across words. Subjects are required to demarcate words within the following type of format: *this is a n example of a coherent sentence*. In the present study the task consists of identifying words within coherent and random versions of a passage written in the target language. In the **Coherent Version** all words form syntactically and semantically acceptable sentences; the sentences form
a meaningful paragraph. The Random Version contains the same words as the coherent version, but the words do not follow the natural order of meaningful text because syntactic and semantic constraints are absent.

Also pertinent to the word-boundary task are the remaining three terms: Context Clueing refers to a pre-task verbal clue intended to provide an expectation for context. Boundary Markers are two hyphens (- -) built into the format of the word-boundary task that appear at the end of a sentence in the coherent versions of the passage, and occur randomly in the random versions. Presentation Sequence refers to the order of presentation of the random and coherent versions of a passage.

Significance of the Problem

Research in the area of comprehension has served to demonstrate the complexity of the reading process. What has become increasingly evident is that a clearer understanding of both the linguistic and cognitive skills involved in reading is necessary to gain insights into this multifaceted process. Studies that neglect one or the other of these aspects may furnish only a superficial view of comprehension. Investigations that provide more complete coverage of these components are relevant for the knowledge they may contribute toward a more comprehensive view of the reading process.

This research has potential value for the empirical support it lends to theoretical models that consider reading in a second-language context an interactive process. In addition, the methodological concern of the study may serve to further establish a means of observing and evaluating the reader's use of context. While the importance and necessity of reading strategies have been emphasized in recent years, tools for measuring
such behaviours are needed to better understand individual differences
in the use of strategies. It is hoped that the present study will provide
insights into one technique designed for the measurement of context
utilization.

Limitations of the Study

This study was partially *ex post facto* in nature, which limited the
ability to draw strong causal inferences. In this respect the interpreta-
tion of results was restricted.

The inability to select subjects randomly from the relatively small
target population was also a factor that may have affected the results
of the study.

The use of the word-boundary task may be restricted to those studies
investigating languages with orthographic systems similar to English.
Moreover, the task may not operate as efficiently with languages such as
German that possess a high frequency of compound words. Findings per-
taining to the use of this instrument were, therefore, generalized only
to Italian and languages similar in orthography.
CHAPTER II
REVIEW OF THE LITERATURE

Introduction

Two major aims of this study were to examine the relationship between reading comprehension and knowledge of a second language, and to measure context utilization by means of a word-boundary task. This chapter is therefore organized into two broad sections. The first reviews theoretical and empirical bases for the examination of specific sources of language knowledge and its use in relation to reading comprehension. Topics in this section include: theoretical views of reading comprehension, the role of prior knowledge in comprehension, sources of knowledge about language and reading, and empirical evidence suggesting the importance of word knowledge, syntactic knowledge, and knowledge of language constraints to reading comprehension. The second section focuses entirely on the questions posed by this research regarding word-boundary task performance. Within this section, previous investigations of word-boundary task performance are reviewed, some methodological aspects of the task are discussed, and the measurement of context utilization is considered in light of the potential validity and limitations of the instrument employed.

Reading Comprehension and Language Proficiency

Current Views of Reading Comprehension

Since its relatively recent promotion to the status of "cognitive
process", reading has aroused the interest of researchers in several specialized areas. Linguists, cognitive psychologists, psycholinguists, and experts in artificial intelligence have brought their own sources of knowledge to bear on reading, not only to arrive at a valid explanation of that process in particular, but also to gain insights into the mental operations basic to all information processing. On the basis of such sources, there remains little doubt that reading comprehension entails highly coordinated cognitive processes. There is less certainty and agreement, however, regarding which processes are involved and how they operate during reading. The lack of consensus concerning the nature of the reading process is hardly surprising considering the formidable questions cognitive matters pose. Up to this time, theorists searching for better explanations of what transpires when a written message is understood, have looked literally in all directions. As a result, at least three distinct types of reading models have evolved; namely, "bottom-up", "top-down", and "interactive". As the terms imply, the models have been classified to distinguish among views of the directionality of the reading comprehension process; the first two types propose opposite processing courses, while the third provides for information flow in both directions.

In general, the ideas represented in "bottom-up" models suggest that reading is primarily a perceptually-oriented activity. The reader registers graphic input then transforms the information into meaning in a series of stages proceeding from features-to-letters-to-words-to-phrases. The stages are ordered (although some may be bypassed) from low to high levels of analysis, therefore, higher-level processes in models espousing
this view have no effect upon lower-level ones. Strictly speaking, comprehension is initiated by sensory or surface structure that must be encoded piecemeal into deeper levels of meaning. Perhaps the most serious proponent of bottom-up theory is Gough (1972). Another example can be found in LaBerge and Samuels (1974), however, a more recent LaBerge-Samuels model (1977) has "top-down" capacity as well.

According to "top-down" models (Neisser, 1967; Levin and Kaplan, 1970; Kolers, 1972; Goodman, 1976) perceptual analyses that begin centrally, within the brain, are guided and modified by the reader's expectations of what will make sense. In this way, reading becomes a problem-solving activity requiring active reader involvement (anticipation, prediction, hypothesis-testing) in the process. Since reading depends to a great extent on the prior knowledge the reader possesses, higher-level processes are predominant; thus, structure is anticipated rather than derived and meaning for the sensory signals is predicted, tested, and identified rather than decoded from the graphic input. Although Smith (1971, 1973, 1975) has not provided a "model" of reading comprehension in the strict sense of the term, his theoretical arguments are characteristic of this top-down perspective.

Rumelhart's "interactive" model (1977) and to a somewhat lesser extent Morton's Logogen Model (1970) are prototypic of a third class of models that involve both bottom-up and top-down modes of processing. In general, an interactive view of the process proposes that a reader is able to construct hypotheses from several independent sources of knowledge simultaneously. These hypotheses, in turn, are confirmed or rejected depending upon their compatibility with the information that is
flowing from the bottom-up and from the top-down at all levels of analysis; there is no fixed point of departure. Consequently, the most likely interpretation of a message is one that satisfies both modes of information processing.

To date, there is no single model of reading comprehension that can be considered complete. The "bottom-up" class, however, has been criticized most often for inability to account for the overwhelming influence of prior knowledge on perception and comprehension. But, empirical evidence of this kind can be explained through top-down and interactive models since both share a basic theory that prior knowledge affects comprehension. In addition, among advocates of this theory of comprehension (regardless of their conflicting views about the directionality of the reading process) agreement can be found concerning the nature of the knowledge sources that readers bring to bear on written text and the strategies readers may employ to integrate this information. The following two sections discuss prior knowledge as the basis for all cognitive processes, in particular, the reading comprehension process.

**Prior Knowledge as the Basis for Comprehension**

Smith's (1975) "theory of the world in the head" and schema theory (Adams and Collins, 1977) explain the role of prior knowledge in comprehension in strikingly similar ways. A fundamental assumption of both is that comprehension is a function of the ways in which an individual's prior knowledge is represented, organized, and used.

Although Smith's (1975) "theory of the world in the head" precedes some of the latest developments in schema theory, it does reflect the earlier work of Lindsay and Norman (1973) and Rumelhart their
collaborator, Collins and Quillian (1969, 1973) and Bartlett (1932), all of whom have combined their efforts in this area. In Smith's view, all prior knowledge about the world is an organized system with three basic characteristics:

(1) a system of categories; (2) sets of rules—or specifications of "distinctive features"—for allocating objects or events to particular categories; and (3) a network of interrelations among the categories themselves... (p. 14)

The categories are abstractions for grouping objects, events, experiences, concepts, and so forth. Each broad category has rules or a "feature list" of distinctive properties for subordinating narrower categories. As a person develops cognitively, new experiences are partitioned; more and more categories are formed. The method of categorization and the allocation of specific objects or experiences to a particular category are highly individualistic decisions, but certainly not arbitrary ones. The system must be meaningful to function for the individual. Categories are therefore related to each other in hierarchical fashion. A subordinate category or "subset" is embedded within a broader category or "superset" which in turn may be a subset of another superordinate category. The interrelationships (or overlap) among cognitive categories are represented in branching tree-like structures or "logical networks" (p. 18). The networks of knowledge do not constitute a complete set of all possible cognitive interrelations. But, where there are gaps in the knowledge structure, the "redundant" or overlapping categories are the "plans" (a term borrowed from Miller, Galanter, and Pribram, 1960) for working out problems or constructing new relationships.
These general plans, according to Smith, allow an individual to know in advance (or, at least, to predict) information that is required to accomplish a desired end. Since prediction "cognitively means the prior elimination of unlikely alternatives" (Smith, 1975, p. 92), plans specify the most probable alternate plans or "routines" for attaining the necessary information. The information provided by each new plan is tested against previous predictions, and evaluated for its appropriateness or inappropriateness in matching the established criteria. Depending upon the outcomes of this analysis of information, original predictions are confirmed or modified. Each plan designates further plans based on its own properties or structural organization. As more details are afforded, uncertainty is reduced and therefore the number of likely candidates among available plans is also reduced. The process of generating predictions, testing them, evaluating the feedback, and confirming or modifying hypotheses continues until matching plans with their predictions results in comprehension. Put another way, the cyclical process repeats itself until the connections among plans (the theory of the world in the head) relate meaningfully to the world outside.

In summary, the theory of the world may be defined as "a system of categories, or discrete units of experience, which are interrelated meaningfully" (p. 111). The interrelations are the core of the theory of the world. As Smith points out:

They enable us to summarize past experience, make sense of the present, and predict the future. Nothing in our lives would be comprehensible if these interrelations were absent (p. 17).

Thus, the theory of the world in the head serves as a summary of all that is already known, as a basis for understanding new information, and
as a source for constructing, confirming or modifying hypotheses. In this way Smith likens the theory of the world to scientific theory. It is also very much like recent accounts of schema theory.

According to Adams and Collins (1977) prior knowledge is represented within "schemata" or classes of concepts. The schemata are arranged hierarchically, with general or broad conceptual categories at the top, subsuming other more specific schemata that may be considered underlying elements, subschemata or "slots" (Minsky, 1975) within a larger schema. Each of the schemata contains descriptions of certain features or aspects compatible with a global representation and also evokes other associated schemata. Each newly embedded schema relays information based on its own properties and specifies additional gaps or slots that must be filled. Therefore, future schemata are elected for their ability to meet the needs established by schemata activated previously. In cases in which the needs are not explicit, values may be assigned momentarily by "default" until a more precise fit can be found among competing schemata. The schemata themselves narrow in scope as they increase in number; thus, those at higher levels are more abstract and general, while lower-level schemata are more concrete and specific. Increasing specificity causes the required details of the relevant schema to become more easily discernible thereby constricting the number of eligible schemata to be accessed. The interrelationships among schemata at all levels constitute a highly constrained system in which prior knowledge is structured and through which comprehension is facilitated. As Adams and Collins explain:
The power of this structure derives from the fact that the top level representation of any schema simultaneously provides an abstraction of and a conceptual frame for all of the particular events that fall within its domain. (p. 6)...The comprehension of a specific situation or story involves the process of instantiation whereby elements in the situation are bound to appropriate slots in the relevant schema. This process not only serves the purpose of filling out the details of the schema, but also of temporarily connecting it to characteristics of the bound schemata (p. 7).

Within schema theory, the knowledge structure described is the basis for all cognitive processes. This theory has received considerable attention and critical examination for its applicability to the reading process. Schema theory "provides a way of integrating our understanding of text with our understanding of the world in general" (Adams and Collins, 1977, p. 41). The same might be said of Smith's (1975) "theory of the world in the head [that] is related to the world around us" (p. 12) except that Smith proceeds to illustrate not how the two are integrated during reading, but how the theory in the head "imposes meaningfulness upon the world..." (p. 12, emphasis added). This slight difference in interpretation manifests itself as a substantial difference when Smith's (1971, 1978) theory of reading and a schema-theoretic view of reading are compared. As alluded to earlier, it is the difference between a unidirectional "top-down" derivation of reading and an "interactive" view incorporating dual processing-controls. From either perspective, the importance of knowledge structure is paramount; however, expectations of how readers use knowledge structure differ. As will be suggested below, both have implications for reading comprehension of a second language.
In the following section specific sources of prior knowledge that have been identified as important to the reading process are discussed within a schema-theoretic model of reading. Additional psycholinguistic theory is also presented describing how the reader's sensitivity to language cues prompts the use of strategies such as prediction, anticipation, and hypothesis-testing.

**Sources of Prior Knowledge and Reading Strategies**

A very natural extension of schema theory to reading is presented in Rumelhart's (1977) "interactive" model. This model predicts that as graphemic strings enter the system, they are registered and stored. Critical features from this "sensory" information are extracted and fed into a "pattern synthesizer" which has available other independent "nonsensory" sources of information: orthographic, lexical, syntactic, and semantic knowledge sources. These knowledge sources are specialized schemata, each containing information about some aspect of the reading process. These schemata must account for all input data, and similarly all sensory information must be synchronized with the activated schemata. To satisfy these requisites, parallel bottom-up and top-down processes are occurring simultaneously at all levels of analysis. The pattern synthesizer integrates all sources of information, both sensory and nonsensory, "to produce a 'most probable interpretation' of the graphemic input... and the reading process is the product of the simultaneous joint application of all of the knowledge sources" (Rumelhart, 1977, p. 588).

Within the pattern synthesizer, all responses generated by knowledge sources are delivered to a "message center" which keeps a "running list of hypotheses" each reflecting statistical probabilities based on
criteria from independent data stores. All initial hypotheses, furnished by high- and low-level schemata alike, are then assessed against one another. The assessment process further establishes grounds for developing future, statistically predictable hypotheses. Essentially, each schema or knowledge source performs a manifold task: it generates its own hypotheses about the input data, scans the message center for hypotheses evoked by other schemata, and evaluates those hypotheses for their relevance to its own knowledge sphere. Based on the evaluation, hypotheses are retained within the communication system or removed. Those that are kept remain tenable until they have been checked out in the same manner by all other relevant schemata. The constraints imposed by the various knowledge sources facilitate this decision-making process.

This formulation of the reading process, although still in its infancy as Rumelhart (1977) himself remarks, is perhaps the most powerful and certainly the most flexible model at the present time. It meets the demands of empirical data (reported in Chapter I) that demonstrate the effects of prior semantic context on reading comprehension. In addition, the interactive model has the capacity to support evidence that letters are perceived more readily when embedded in familiar words (Reicher, 1969; Wheeler, 1970) and word perception depends on the syntactic environment (Tulving and Gold, 1963; Kolers, 1970; Weber, 1970; Stevens and Rumelhart, 1975) as well as the semantic environment (Schvaneveldt, Meyer, and Becker, 1976; Tweedy, Lapinski, and Schvaneveldt, 1977). Also explained within this model are research findings suggesting that grammatically and semantically acceptable sentences are processed with greater facility than anomalous sentences (Miller and Isard, 1963;
Martin, 1968; Slamecka, 1969) and syntactic and semantic constraints enhance lexical decisions (Klein and Klein, 1973). This literature suggests that, in principle and in practice, reading comprehension depends on the reader's ability to interrelate various sources of knowledge about language and the world.

According to Rumelhart, independent knowledge sources contribute concurrently to the interpretation of input. Each source can suggest, and at the same time, limit, and strengthen or weaken the hypotheses coming from other knowledge sources. Knowledge of orthographic structure, including featural knowledge, letter-level knowledge, and letter-cluster knowledge, offers possibilities and probabilities that certain lexical items will be perceived. Lexical knowledge rates the information supplied by orthographic knowledge in light of the frequency of occurrence of such letter fonts, and posits its own logical hypotheses based on that data and syntactic and semantic knowledge available at the time. Lexical knowledge can also shape the nature of hypotheses to be generated by sources of knowledge below its level (orthographic) and above (syntactic and semantic knowledge). Syntactic knowledge provides information about probable word sequences that fit the information in its own sphere, as well as data reported from orthographic, lexical and semantic knowledge sources. Syntactic knowledge functions either to initiate and/or confirm lower- and higher-level hypotheses alike. Finally, the plausibility of hypotheses developed with respect to the semantic content of a message must be compatible with prior semantic knowledge and must fit the information specified by syntactic and lexical knowledge sources. In turn,
semantic knowledge can guide and modify hypotheses formulated by all other knowledge sources that are operating contemporaneously and collectively.

While the potential value of all knowledge sources working in cooperation with one another is recognizable at any point in the reading process, the extent to which each source is brought into play during a particular reading probably depends on the needs of the reader and/or the difficulty of the text itself. If the material is problematic due to lack of subject-matter knowledge, the reader will necessarily make greater use of orthographic, lexical, and syntactic knowledge. On the other hand, readers encountering difficulties caused by insufficient knowledge of a language may require more semantic context. In the latter case, with only fragmentary knowledge of the language (i.e. some knowledge of the orthography, words and grammar) and a reasonable amount of pragmatic information, it is possible that comprehension will nevertheless occur. Under both of the conditions mentioned, however, the reading process may necessitate highly developed processing strategies.

Rumelhart's representation of the cognitive system in action is also a description of the strategic processing that occurs during reading. Knowledge sources applied purposefully at all levels of analysis constrain the flow of information thereby allowing the reader to make predictions, test, verify or modify hypotheses, and create expectancies for future information. It may be that processing strategies bridge the gaps among knowledge sources. If so, readers may be able to comprehend discourse with less knowledge than might be expected from any one source; proficiencies in one area would compensate for deficiencies in another.
The interactive model of reading would account for compensatory mechanisms (Stanovich, 1980; Schwartz, 1980) through strategic processing capabilities.

Although the consequences of strategic processing are just beginning to be understood in the context of the recent interactive model of reading (Stanovich, 1980; Schwartz, 1980), strategies have been discussed in the literature for some time primarily by "top-down" theorists. Strategies are, in fact, the essence of such theory, the assumption being that information in reading is redundant. Sources of knowledge about the orthographic, syntactic, and semantic systems of language, as well as prior contextual or world knowledge overlap and thus benefit each other by simultaneously supplying cues to the same bit of information. The greater the redundancy or overlap of "nonvisual" information, the greater the predictability of a message, and the less "visual" information is needed to identify meaning (Smith, 1978). Readers who capitalize on regularities and the probabilities of likely occurrences of linguistic elements within the context are able to sample the text (foregoing many of its redundant cues), select the most informative parts, make sophisticated predictions, and anticipate information along the way. Strategic behaviour such as this is not mere theoretical speculation, but an empirical reality appearing in many different areas of study.

Research relating to lexical ambiguity has reported that prior contextual knowledge allows subjects to judge in advance (or predict) the correct meaning of ambiguous words (Schvaneveldt, Meyer, and Becker, 1976; Tweedy, Lapinski, and Schvaneveldt, 1977; Swinney and Hakes, 1976). Several eye-voice span studies have found shorter EVS's for lists of
random words (Gibson and Levin, 1975) as well as for high content (less predictable) sentences (Wanat, 1976), and longer EVS's for more constrained sentences (Levin and Kaplan, 1970). Since the eye-voice span is the number of words that are correctly reported beyond the point at which the visual stimulus is removed, these findings suggest that greater constraints create expectancies for upcoming words. An alternate explanation of these observations is that the eyes are always some distance ahead of the voice, thus the EVS is a report of words remembered. If so, the advantage of more highly constrained sentences is less a function of predictability than of meaning. Wanat's study, however, showed longer fixation times for less predictable sentences, and in particular for high content words such as verbs, ruling out perhaps the second explanation for his findings at least.

Analyses of oral reading errors also suggest that predictive or anticipatory behaviour is prompted by semantic and syntactic structure. Weber (1970) found that 90% of errors made by first-grade readers did not violate the grammatical class of the word substituted. Similarly, with adult readers, approximately 70% of the substitution errors were the same part of speech as the correct word (Kolers, 1970). Two other types of oral reading errors reported by Kolers further demonstrate reliance on semantic and syntactic constraints to anticipate text. Both types were made by bilinguals when presented with materials mixed with English and French words forming coherent text (eg. "Son cheval, suivi by two hounds...", vs. "His horse, followed de deux bassets..."). Despite the language mix, native speakers of either language often continued to read the text as if it were written in the same language. They tended to
"translate" into the language in which they were processing at the time, thus ignoring the visual input and substituting their anticipations. Similarly, the same subjects committed order errors. When materials contained distortions of syntax (for example, "made resound the earth") several subjects (percentages were not given) automatically corrected the disorder which appeared to cause little distraction or interference to them. Both "translation" and "order" errors went unnoticed by many subjects, since few corrected these errors.

The tendency to anticipate or predict grammatical form was also shown in a study of oral reading errors conducted by Stevens and Rumelhart (1975). In a close experiment, 98% of substitution errors were grammatically acceptable and about 80% of the time, the predicted restorations of text were consistent with the syntactic class of the deleted words. Another experiment by Wisher (1976) manipulating syntactic and semantic formats of stories demonstrated that when sentences followed a logical sequence in terms of semantic content, processing time decreased. In addition, when expectations of fixed syntactic patterns were induced, subjects used prior knowledge of grammatical form to anticipate later sequences. Thus, both prior syntactic and semantic knowledge can be utilized effectively for creating expectancies in the reader.

Cloze task performance itself, and similarly word-boundary task performance (see Wildman and Kling, 1979) also tap several strategies such as anticipation, prediction, and inference to generate expectancies about upcoming words. Moreover, they can reflect strategies employed by the performer to relate the natural sequences of linguistic elements with appropriate extralinguistic contexts. According to Oller (1979)
such tasks invoke and challenge the efficiency of an individual's "expectancy grammar" that incorporates "the language user's knowledge of how to map utterances pragmatically onto contexts outside of language and vice versa (that is, how to map contexts onto utterances)..." (p. 24).

Through his notions of an expectancy grammar, Oller argues firmly that language tests must reflect the use of both linguistic and extralinguistic constraints to integrate new information with prior knowledge. He therefore develops a strong case in favour of

...any procedure or task that causes the learner to process sequences of elements in a language that conform to the normal contextual constraints of that language and which requires the learner to relate sequences of linguistic elements via pragmatic mappings to extralinguistic context (p. 38).

Nearly all of the research cited above could be taken as evidence of the reader's use of strategies for either or both of two purposes stated above: i.e. to facilitate ongoing processing of text and to integrate new information with prior knowledge. While the use of strategies for both purposes is supported by top-down or interactive assumptions, strategies to speed ongoing word or sentence processing are obligatory from the former viewpoint but only optional from the latter. This distinction, although theoretically real, is not easily observable for two related reasons. First, readers rely on redundancy or statistical predictability to engage in strategies for both of the purposes mentioned. Secondly, redundancy, the construct central to many top-down views of the reading process is the same as statistical predictability that facilitates operations within an interactive system. As Adams(1979) explains: "...any variable that alters the redundancy and, by implication, the perceptibility of a stimulus... must also alter its
predictability or a priori expectancy..." (p. 140). Consequently, it is difficult to separate the two theories empirically. One exception in this respect, however, is found in the body of literature addressing comprehension differences between good and poor readers.

By top-down standards, good readers comprehend better because their use of redundancy (knowledge of orthographic, syntactic, and semantic constraints) frees them from processing low-level information thus allowing more time to make high-level predictions that guide perceptions at lower levels. Poor comprehenders, who lack the ability to use contextual constraints, process language at lower levels and thus lose the benefits of high-level conceptual processes. By contrast, within an interactive framework, the differences between good and poor readers would be attributed to varying amounts of stored knowledge and/or the inability to integrate prior knowledge with new information contained within a text. Once again, there seems to be little disparity between top-down and interactive explanations since both are drawing the distinction between knowledge the reader has and the knowledge she or he must use to comprehend a given text. In the first instance, however, poor readers are at a greater disadvantage since according to top-down theory they are doomed to low-level processing. On the other hand, the mutual processing of both high- and low-level information provided by an interactive reading model lifts these restrictions from poor readers giving them free rein to rely on either mode.

Empirical evidence suggests that poor readers are just as reliant on contextual constraints (Weber, 1970; Kolers, 1975; Allington and Strange, 1977) occasionally even more reliant (Allington and Fleming,
1978; Schvaneveldt, Ackerman, and Semlear, 1977; West and Stanovich, 1978) than good readers. In a recent article, Stanovich (1980) argues that these data are only understandable within an "interactive-compensatory" model of individual differences in reading. As Stanovich points out, good readers may be more sensitive to contextual constraints (Isakson and Miller, 1976; Katz and Wicklund, 1972; Perfetti and Roth, 1981), display a greater knowledge of language structure or cueing systems (Lott and Smith, 1970; Weinstein and Rabinovitch, 1971; Clay and Imlach, 1971; Allington, 1978) and employ them to a greater extent in problem-solving tasks (Ruddell, 1965; Klein, Klein, and Bertino, 1974), but, they may not be as dependent upon contextual constraints as poor readers in an actual reading situation. It may in fact be the overreliance on contextual information by poor readers rather than the use of constraints by good readers that distinguishes between their comprehension abilities. The former may rely on context for accuracy or to compensate for slow, non-automatic word recognition skills, while the latter may be reliant on context for speed (Allington, 1978; Perfetti and Hogaboam, 1965). The implication is that an overuse or misuse of context can actually drain the poor readers' cognitive capacity for integrating new, incoming information into their own knowledge structures. Conversely, the more effective use of context by good readers to activate information stored in operational memory accounts for better comprehension (Posner and Snyder, 1975). In favour of this position is empirical evidence suggesting that good readers exercise more strategic control over cognitive processes to integrate new and given information (Guthrie,
Cromer et al. (1971, 1967) have noted that comprehension difficulties may be a result of a "mismatch" between a reader's organization of linguistic patterns and his or her cognitive-structural organization. Investigations with two distinct groups of poor readers, identified as "deficit" (deficiencies in vocabulary and word identification) and "difference" (poor comprehension but adequate vocabulary skills) have shown that readers skilled in word identification still suffer comprehension breakdowns due to the manner in which they organize input. Good readers, on the other hand, appear to impose order and organize input into meaningful patterns.

Guthrie (1973) concluded from a study of relationships between reading sub-skills and young readers' comprehension scores that many reading difficulties stem from failure to utilize a variety of potential sources of information. Schwartz (1977) interpreted Guthrie's findings to reflect a general "strategic deficit" that manifests itself at a number of different levels in comprehension. Schwartz implies that although general reading comprehension problems may be related to knowledge or conceptual deficits, the inability to integrate various reading sub-skills due to organizational or strategic deficits may be a more critical factor.

Bearing in mind the arguments presented in this section, it is reasonable to assume that successful comprehension is a function of the reader's knowledge structure and use of this structure for constructing meaning from text. Proportionately, language knowledge and its use is
only a segment of the wealth of prior knowledge operating in reading, but a significant portion particularly if the text is written in non-native language. For this reason, factors implicated in a learner's second-language proficiency are implicated also in reading comprehension of that language. The following section reviews literature in the areas of first and second language that affirms the importance of knowledge factors about language to reading comprehension.

**Word Knowledge, Syntactic Knowledge, and Inference**

Identifying the factors involved in reading comprehension has been the objective of many factor-analytic studies. In the first study of this kind, Davis (1941) was able to reduce nine testable reading skills to five significant factors which he named in order of importance: knowledge of words meanings, verbal reasoning, sensitivity to implications, following the structure of a passage, and recognizing the literary techniques of a writer. Subsequent reanalysis of Davis' 1941 data by Thurstone (1946) resulted in further reduction of these factors to three, renamed: vocabulary knowledge, ability to draw inferences from a paragraph, and ability to grasp the main idea of a paragraph. Further empirical research by Davis (1944) led him to conclude that "word knowledge and reasoning in reading account for virtually all of the variance in reading comprehension" (Davis, 1968, p. 508). In 1968, Davis reconfirmed this statement in light of new data analyses which found 32% of the variance in eight reading skills to be unique to "memory for word meanings", and 20% of the non-error variance to be attributable to "drawing inferences about the content of the materials read" (p. 542-43). At that time, Davis further defined "reasoning in reading" as "a combination of weaving ideas together and drawing inferences from them" (p. 544).
Other factor-analytic studies have uncovered reading comprehension factors with loadings ranging from .41 to .93 for vocabulary and reasoning abilities as well (Fruchter, 1948; Botzum, 1951; Wrigley, Saunders, and Newhaus, 1958; Clark, 1972). In addition, two research teams working separately (Brinton and Danielson, 1958 and Stolurow and Newman, 1959) performed factor analyses on a matrix of correlations between internal material factors of reading passages and comprehension scores, originally prepared by Gray and Leary (1935). Both interpreted two principle factors to be a vocabulary factor and a sentence factor. Authors of the 1959 study specified that "relative difficulty of words" accounted for 34% of the variance in comprehension and "relative sentence difficulty" contributed 20% to the total variance.

The influence of these latter two studies is apparent from the fact that nearly all readability formulas take into consideration some measure of vocabulary load and sentence length or syntactic complexity (Chall, 1958; Klare, 1963). Both of these researchers acknowledge that some measure of word frequency, familiarity, or difficulty has more validity than any other single factor as a predictor of text difficulty. Indices of both word and sentence or syntactic complexity, however, are necessary to predict difficulty sufficiently. Coleman (1971) studied reading difficulty in terms of morphological, syntactic, and semantic factors. He found syntactic complexity to be an important factor, but measures of word complexity alone explained approximately 80% of the variance. Bormuth (1966) confirmed that vocabulary load and grammatical complexity are significantly related to reading difficulty, but also advocated a new approach to estimating readability.
This new approach was the cloze procedure. As a readability tool it has demonstrated greater reliability, predictive and concurrent validity than any combination of factors in the past. It is also the most highly correlated with reading comprehension and with a host of other cognitive and verbal abilities. Oiller and Perkins, (1978, 1980) report that when cloze tests have been factor analyzed with batteries of language tests, a single "global language proficiency" factor has emerged consistently. A factor analysis of 8 cloze tests and 18 cognitive tests conducted by Weaver and Kingston (1963) found cloze tests to be most closely related to a factor labelled "redundancy utilization". Bormuth (1967) found correlations ranging from .73 to .84 between cloze tests and multiple-choice comprehension tests measuring vocabulary, explicitly stated facts, sequences of events, inferences, causal relationships, main ideas, and authors' motive. Judging from this list of measures a plausible explanation of the cloze task's robustness is that it reflects "reasoning" or the ability to "weave ideas together and draw inferences from them". Several researchers who have found strong associations in the range of .52 to .76 between the cloze and both verbal and nonverbal reasoning tests share this opinion (see Rankin, 1965).

The studies presented above clearly demonstrate that reading comprehension "is not a unitary mental skill or operation" (Davis, 1968, p. 542). At least three factors have shown some unique variance associated with that construct. Among these, the most salient appears to be word knowledge, followed by reasoning or "a combination of weaving ideas together and drawing inferences from them". The third factor, sentence structure or complexity, is a material factor. By deduction, however,
it can also be considered in terms of the reader to imply a logical or meaningful blend of syntactic and semantic knowledge for understanding sentences. Stated in this way, it is difficult not to notice the correspondence of this factor with "reasoning". The two factors may, indeed, have received the same name from their founders, if not for the fact that they were derived from different criteria. From the vantage point afforded by the theories discussed earlier in this chapter, it is clear that a great deal of reasoning is required to integrate knowledge about how language works with knowledge about the world. Vocabulary measures that are so highly correlated with measures of verbal and non-verbal intelligence may be tapping, in reality, the global construct "world knowledge".

Much of what readers know about their own native language may be inextricably linked to their knowledge of the world. It has been suggested, for example, that one reason word knowledge is so essential to reading (Pike, 1977; Spache, 1962; and Stotsky, 1976) is because it may be a "reflection of the extent of exposure to the culture... that is crucial for text understanding" (Anderson and Freebody, 1979). Similarly, syntactic knowledge cannot be divorced entirely from other aspects of an individual's background knowledge including how objects are related or events and actions are sequenced temporally and spatially. Linguistic and extralinguistic patterns are established through the same structures that carry meaning (Lefevre, 1964; Fries, 1963). Studies with both natives and non-natives of English have shown that children acquire language as a result of direct interaction with their environment and that a strong link exists between syntactic maturity and reading comprehension
ability (Ervin-Tripp, 1974; see also King, Holley, and Weber, 1975).
In addition, syntactic knowledge gives the reader an intuitive sense for judging the grammaticality of certain word sequences and disentangling ambiguities (Chomsky, 1965; Miller, 1973; Kolers, 1970; Miller and Isard, 1963).

A similar overlap, however, between the reader's world knowledge and knowledge of a non-native language that is acquired in a classroom, apart from other cognitive development, cannot be assumed. The second-language learner's task, in fact, is to make such connections between prior knowledge and the novel language system and to construct new knowledge bases when direct relationships cannot be found. Granted, the associations a learner discovers or develops will vary depending upon cognitive and affective factors as well, but all second-language learners must choose some method of categorizing new lexical items and specifying new rules for rendering word sequences meaningful. To the extent that these methods are successful, the learner should begin to see patterns or redundancies. During the incipient stages, however, the learner's knowledge of the second language is still, at best, also rudimentary.

Self-reports of students of English with diverse native-language backgrounds have indicated that once individuals feel they can handle the basic grammatical rules of the new language system, a lack of word knowledge is considered the greatest obstacle to comprehension (Yorio, 1971). It should be noted that these reports came from foreigners studying English in the United States, therefore, American students, for example, who are learning a second language in an American classroom, may find inadequate lexical knowledge to be an even greater hindrance to comprehension. Lexical knowledge is a level of representation at which
semantic and phonological aspects of all languages are joined. Native speakers of a language have an "intuitive" command of the information about linguistic forms that is represented at the lexical level (Chomsky, 1970).

In a study of second-language readers' comprehension processes, Phillips (1974) also recognized the problem of insufficient vocabulary and observed that in many cases readers ignore tense markers. In addition, she identified subtasks that some readers undertake to enhance comprehension. These included categorizing words grammatically and recognizing cognates and root words. An interesting and important point made by Phillips is that the less proficient reader may not recognize readily what by traditional standards might be termed a "cognate". This observation reinforces the idea that methods of categorization are highly personalized decisions and some readers will have better-organized and more easily accessible knowledge structures for processing the new language than others.

Acknowledging the handicaps second-language readers must overcome in order to comprehend discourse, Cates and Swaffar (1979) pointed out that recent studies have provided valuable but "incomplete" knowledge about what reading comprehension actually entails. They maintain, however, that:

Reading for comprehension is a reasonable objective when learners know strategies for prediction and inference... foreign language students should be required to rely on prediction and inference to augment their limited knowledge of the language (p. 5).

To a great extent the findings of Neufeld (1980) substantiated these remarks. In a study of the effects of new-word density on reading comprehension, she found that fluent second-language readers of English
employed strategies of prediction and inference to restore context to
cloze passages. Neufeld stated that the fluent reader derives meaning
"more through the use of skills involving knowledge of the grammar and
knowledge of the world than through the decoding of the visual marks on
the page" (p. 98). She concluded that "It is these strategies, rather
than the possession of a large vocabulary... that denote success"
(p. 110). Neufeld further observed that students who lack knowledge of
syntax and are unaware of language redundancies, are also unable to make
accurate predictions.

Neufeld's study also showed that high-proficiency students have a
greater tolerance for unstructured material than do less proficient
students. More specifically, two groups of subjects were given a pas­sage to read in which the words of each sentence were scrambled, but the
order of sentences and paragraphs was not manipulated. While students
proficient in the language appeared to attack the reading with problem­solving strategies, creating a context that was virtually absent, less
proficient students were reluctant even to attempt the reading of this
passage. Similar hesitations, however, were not found among subjects of
either low- or high-proficiency groups who received the same passage, the
appearance and content of which were equally mutilated, if not more so.
One difference between the scrambled passage condition and all other
treatment passages was that the natural order of words within sentences
was destroyed. All other treatment groups read passages in which origin­
al English words had been replaced by nonsense words, but the order of
words and sentences was that of the original passage. Subjects in these
latter groups, regardless of proficiency level, tolerated high ratios of
nonsense to English words and appeared to behave in a strategic manner.
Only low-proficiency students in the scrambled text condition found the passage impossible to handle.

There appears to be a paradox in this observation. Less skilled readers were able to derive meaning from "nonsense" that was structured, but when syntactic order was not maintained in a text of potentially meaningful words, unskilled readers perceived nonsense. There is some suggestion in these data that less proficient readers see "nonsense" in disorder and are therefore sensitive to syntactic constraints. Moreover, they may depend on such constraints for ongoing text processing. When they "perceive" the absence of sequential redundancies, unskilled readers may feel there is no information upon which to rely for comprehension. This interpretation would be consistent with first-language studies that have shown poor readers to be more reliant on contextual constraints for processing incoming words than good readers. Also consistent with first-language reports is the behaviour of the skilled second-language readers in Neufeld's study who put together a context despite the disarray of scrambled text. Perhaps the skilled second-language reader, like the good reader of native language, utilizes context primarily to integrate new information with prior knowledge rather than to facilitate the identification of words—that is, has a set of strategies for imposing order on varied kinds of information.

If these extrapolations are accurate, second-language students will behave in a manner similar to native-language readers depending upon their knowledge of the target language. Furthermore, successful comprehension will be dependent upon the readers' ability to use strategies to compensate for insufficient knowledge of words and language rules. Despite the fact that readers of all proficiency levels will use their knowledge
of contextual constraints to derive meaning, the efficiency of their strategies is directly related to their proficiencies. Thus, the predictive strategies of unskilled readers may still be inadequate for conceptual processing to occur.

On the one hand, all readers of a second-language need the advantage of both bottom-up and top-down modes of processing that an interactive model of reading furnishes. On the other hand, the actual behaviour of many readers may be more characteristic of top-down assumptions. More specifically, second-language readers must use context not only to integrate information, but also to recognize upcoming words and establish patterns. Their unfamiliarity with the language requires that they use context for both purposes to a greater extent than either good or poor readers of native language. If readers can predict incoming information with a minimum of low-level information, there will be more capacity remaining for conceptual processes. Moreover, unless readers can use context to specify patterns of incoming words readily, the readers' frustration level may become so great that conceptual processing is no longer a possibility.

In concluding this discussion, it is apparent that reading comprehension of a second language can be understood only within an interactive and/or top-down framework. Top-down theory emphasizes the facilitating effects of using contextual constraints to identify meaning and proceed directly to higher levels—a process that may be necessary for readers lacking linguistic knowledge. In addition, an interactive view characterizes the flexibility of readers more richly than previous theories, by identifying the use of all knowledge sources simultaneously. A more valid model of second-language reading comprehension is afforded by an integrative theory.
Summary and Overview

There is strong agreement among theorists that word and syntactic knowledge as well as the readers' ability to predict and infer meaning are essential factors influencing comprehension. These same factors apply also in the area of second-language research. Moreover, in that field, there is a keen awareness of the fact that due to limited knowledge of a second language, readers must engage more actively in strategic processing than must readers of a native language. To be able to employ strategies, they must rely on sources of knowledge other than their knowledge of the second language. They must, in fact, draw upon all available information including knowledge of their own language, world knowledge, relevant contextual knowledge, and so forth at every stage of processing to facilitate ongoing text processes as well as integrative processes. The use of strategies for both purposes is, perhaps, obligatory in many instances for comprehension of a second language to occur.

One purpose of this study was to investigate the second-language readers' use of strategies for prediction and inference. It has been hypothesized that readers will rely on knowledge of contextual constraints to utilize such strategies. A task designed to tap the readers' use of strategies, and by implication, their knowledge of constraints, offers possibilities for examining factors that may affect strategic processing, and for measuring the readers' use of context. The issues relevant to the use of a word-boundary task for both of these purposes are the topics of discussion in the following section.
Word-Boundary Task Experimentation

Methodological Aspects

A task designed to examine the use of contextual cues for visual word identification in a continuous processing situation, was developed by Klein and Klein (1973). This instrument, known as the word-boundary task, is prepared such that letters comprising individual words are printed uniformly one space apart. Across words, no additional spaces are provided, therefore subjects must locate words as quickly and accurately as possible by placing a slash between the last and first letters of adjacent words.

In the simplest form in which the WBT has been employed, subjects respond to two different forms, one in which words are presented within a coherent context, the other in which words are arranged randomly. Performance on the two tasks, defined by the number of correctly identified words and/or the number of errors, is compared to determine the effect of context (present in coherently organized material) on word identification.

All of the research relevant to WBT performance has been experimental and has been conducted in English with native English speakers. Many studies in which the WBT has been used, however, have reported the use of contextual information by children (Klein, Klein, and Bertino, 1974), differential use of contextual cues by normal and disabled readers (Klein, 1977), and developmental increases in context utilization\(^1\) from third grade to college-aged subjects (Klein, Klein and Doris, 1973; Goldsmith, 1974; Klein, 1976). This previous use of the WBT with various

\(^1\)For differences of opinion concerning the interpretation of these findings see Stanovich, 1980 and Schwartz, 1980.
age and ability groups suggests that the instrument may be employed ef-
fectively with students of a second language, provided the language has
an orthographic system similar to English.

An investigation of WBT performance with second-language learners
of Italian was undertaken by this research. An experiment was designed
to examine the effects of coherent versus random text organization on
subjects' ability to identify words within a word-boundary paradigm. In
addition, verbal contextual clueing and sentential boundary markers were
examined for their main and interactive effects. Each of these factors
has been explored in a similar fashion by researchers with first-language
subjects. The studies that have addressed these variables in the context
of a WBT are discussed in the following order: effects of coherence on
task performance, boundary marker effects, and contextual clueing effects.

Effects of Coherence on Task Performance: Citing previous research
that had demonstrated the facilitative effects of syntactic and semantic
constraints on word perception (Miller and Isard, 1963; Martin, 1968;
Slamecka, 1969), Klein and Klein (1973) sought to examine the use of lin-
guistic cues in word identification. They conducted two experiments with
undergraduate college students using a WBT. For the first experiment,
four passages were prepared in a WBT format, one of rather simple prose
and the other three constructed at sixth-, third-, and first-order ap-
proximations to English respectively. Thus, the materials varied from
the logical sequence of each word in the prose passage, to random strings
of words in the first-order approximation. All subjects performed on
each passage under paced time conditions. Only the presentation order
of passage was varied to control for sequence effects. There were,
however, no significant order effects, and the data were pooled for the
analysis of differences among the four passages varying in probabilistic
structure. Data were scored for the number of correctly identified words
and for the number of errors. The results of separate analyses performed
on each measure were consistent. Performance on the WBT increased signi-
ficantly with greater approximations to coherent text. Only comparisons
between the third- and sixth-order approximations were non-significant.

The second experiment was designed to explore the use of syntactic
and semantic constraints in more detail. Grammatical, ungrammatical, and
anomalous passages were therefore constructed, and the same procedures
and design used in the first experiment, were applied. Again, no order
effects were reported and the results of this experiment were similar to
the former. When the dependent measure was the number of words correct,
performance on the grammatical form was significantly higher than on the
other two forms. Similarly, with the error data, significantly greater
numbers of errors were found for the ungrammatical form than for gram-
matical and anomalous forms; differences between the latter two forms,
however, were non-significant.

Klein and Klein concluded that in both experiments the accuracy and
speed of word identification was enhanced by the use of both syntactic
and semantic cues. These same conclusions were reached in a later study
(Klein, Klein, and Bertino, 1974) conducted with children who performed
on coherent and random versions of passages prepared in the WBT format.

In both of the experiments described above, subjects were given
several passages with varying amounts of context. In contrast, this
study employed only one passage written in the target language. Each
subject, however, received both the coherent passage, as well as its randomized version. The sequence of the two tasks was also controlled to guard against extraneous effects due to that variable.

**Effects of Boundary Markers:** Klein, Klein, and Hildum (1974) conducted a series of experiments based upon "active" information processing (Kolers, 1970; Neisser, 1967; Smith, 1971) and empirical studies indicating that smaller linguistic units are more easily recognized when embedded within larger units (Brown and Hildum, 1956; Ladefoged and Broadbent, 1960) and constituent and articulatory phrase units reflect the underlying syntactic structure of a sentence (Fodor and Bever, 1965; Garrett, Bever, and Fodor, 1966; Graf and Torrey, 1966). Using WBT performance as the criterion measure, three experiments were designed specifically to compare the effects of organizational structure provided by these phrase units as well as sentential boundaries on the processing of written language. For each experiment the WBT format was altered to provide phrase and sentence information by the inclusion of two hyphens (− −) to represent breaks corresponding to articulatory phrase units, constituent phrase units, and/or sentence boundaries.

In the first of these experiments, subjects were assigned to an articulatory-phrase condition or a constituent-phrase condition. Six passages with a mean sentence length of approximately 19 words, were prepared in the WBT format. For each condition, the appropriate phrase unit (articulatory or constituent) was marked by (− −) in three of the passages, and the placement of hyphens in the remaining three passages was random (the random passages served as controls). Subjects in each condition therefore had six trials of WBT performance. The sequence of
passages was controlled, but no effects were found for that variable. Data, scored for the mean number of correct words across phrase and control trials, were therefore pooled. The results showed that WBT performance was facilitated by the presence of phrase-unit information. No substantial differences were found however, for the type of phrase-unit information provided. According to omega squared values calculated for the two phrase conditions, each accounted for about 16% of the variance. The WBT forms for articulatory and constituent phrase cues, however, had a small percentage of breaks in common that corresponded to sentence boundaries. The suspicion that phrase and sentence information may have been confounded led to another experiment.

In this experiment the same six passages from the first experiment were used in generating three forms in which sentential boundaries were marked, and three control forms in which breaks occurred within sentences, but not at phrase units. No significant differences were found for the sentence cues, implying that the effects of phrase units in Experiment I were not simply a function of the overlap of sentential information in those forms.

The third experiment in this sequence was designed to examine sentence cues in shorter sentences. Four passages averaging six words per sentence, as opposed to 19 in the first two experiments, were prepared in the WBT format. Each subject received two sentence and two control trials. Data were scored and analyzed as explained for the first experiment, and the difference between sentence and control forms was highly significant. Word identification was facilitated by the use of sentence information. Moreover, sentence cues accounted for over 37% of the variance indicating a greater contribution to performance than was found with
articulatory and constituent phrase units in Experiment I, even though there were greater distances between breaks on the sentence forms (six words on the average as opposed to four words between articulatory or constituent phrase units). The researchers offered two explanations for the greater effects of sentence cues, arguing first that sentences are "less arbitrary" or less dialect-specific than phrases. Secondly, sentence information might demonstrate greater effects than any single type of phrase unit, since, in short sentences the "sentence breaks are both pause points and major constituent boundaries" (p. 340).

The researchers concluded from their findings that although performance on the WBT was facilitated by phrase structure information, "sentence boundary information can provide an even stronger facilitating effect than phrase cues" (p. 340). On the one hand, breaks at phrase and sentence boundaries served to organize the material into smaller, meaningful units reducing the amount of information between breaks and increasing the predictability of sequences. On the other hand, while "control" breaks, randomly distributed constituted a "free slash", they may have disrupted the flow of information, interfered with motor and/or attentional sets, and may have also placed greater demands on memory. These factors may have decreased performance on the control forms of the WBT.

The experiments just described raised questions about the effects of boundary markers on random and coherent WBT performance. In addition, they demonstrated the adaptability of the WBT for examining contextual features within sentences.
Effects of Contextual Clueing: Schwartz (1977) acknowledged the evidence that overlapping contextual cues within sentences enhanced word identification on the WBT, and further reasoned that prior "thematic context" might facilitate this process in a similar manner. He therefore tested this hypothesis in two experiments. In the first, all subjects performed on random and coherent passages prepared in a WBT format. Only half of the subjects, however, were given a framework statement. This statement, placed above the passage, was to be read by subjects prior to performing the word identification task. Analysis of the data found no significant effects for the presence or absence of a context statement. Schwartz suspected, however, that subjects had been able to construct a context on their own for the passage used in this experiment. He therefore asked whether an effect might be found for a framework clue with passages less likely to be interpreted spontaneously. This question was addressed in a second experiment.

In this experiment random and coherent versions of the well known "If the balloons popped..." passage developed by Bransford and Johnson (1972) were prepared in the WBT format. This time the contextual framework was provided in the form of a picture (designed also by Bransford and Johnson) that depicts in detail the situation described in the passage. Subjects were assigned to random or coherent task conditions and framework-no framework conditions. No interaction was found between the framework and coherence factors, but significant main effects for both factors indicated that coherent passage organization and contextual information at the framework level enhanced word identification.
As Schwartz pointed out, the differential effectiveness of the framework clues provided in these two experiments was not necessarily a function of the comprehensibility of the passages, but may also have been due to variations in the manner and duration of the framework presentation. Whereas the framework statement of the first experiment was presented at the time of the word identification task, the pictorial framework was projected for 30 seconds prior to beginning the word identification exercise.

The decision to include contextual clueing as a variable in the design of this study was made prior to reviewing Schwartz's experiments. This was decided primarily because of the intention to examine the WBT with second-language subjects who have a very limited knowledge of the new language. Although Klein's et al. studies had shown that native-language subjects discover and use context when present, the question arose as to whether or not second-language subjects would display the same abilities during the very brief time period allotted for performing the task. Regardless, even if subjects could discover context, the discovery would take additional time that might alter the results. The experiment was designed, therefore, to test the effects of verbal contextual clueing on WBT performance. Despite the fact that Schwartz found no effects for the verbal statement in his first experiment, he explained that the duration and presentation of this stimulus may have affected the results. Furthermore, other researchers have found facilitating effects for titles of ambiguous or obscure passages in the areas of both first and second language (Anderson, Reynolds, Schallert, and Goetz, 1977; Adams, 1981). In agreement with Schwartz's remarks, the contextual
clue provided in this study was presented on a separate cover sheet prior to the WBT; subjects were allowed ample time to read this brief statement (written in English, of course!).

**Theoretical Explanations:** The research described in the previous paragraphs offers strong evidence that context utilization, based on knowledge of linguistic constraints, is reflected in WBT performance. To explain the findings of these experiments, it is necessary to assume some top-down mode of information processing, since information at higher levels enhances the identification of patterns at lower levels. These effects are easily explained by top-down theory, since the basis of such theory is that conceptual level information guides lower-level perceptions. In addition, the effects are equally accountable in terms of an interactive reading model that postulates the joint application of independent knowledge sources at all levels of analysis. From this standpoint, prior syntactic and semantic knowledge sources create expectations for identifying lexical items; hypotheses generated at higher levels converge with, and are verified by the information flowing from the bottom, upward.

Both of these theories provided bases for the investigation of the WBT as it was used in this study, however, other findings that have emerged from research with this instrument suggest that an interactive model has greater explanatory value. More specifically, in two experiments in which secondary tasks were imposed on subjects performing a WBT, differential effects of context utilization were found to be more supportive of interactive assumptions. In one of these experiments, Schwartz (1977) prepared random and coherent passages in normal and reversed orthographic formats of the WBT. The reverse-order condition was
created such that letter features were unchanged, but subjects had to follow letter sequences from right to left as opposed to the normal scanning pattern. This produced a condition that increased the conventional disruptive effects of the WBT for word identification. The results indicated that coherent organization of text enhanced performance, reverse orthography reduced performance, and the two variables interacted significantly. Viewing performance on the coherent task relative to base rate performance on the random task, Schwartz interpreted the interaction to signify "...a marginally significant \( p < .05 \) tendency for coherence to increase performance more for the reversed orthography than for the normal orthographic condition..." (p. 10). According to Schwartz, the greater "relative" facilitative effect of context for the reversed format indicated that high-level information increased in relation to the value of lower-level codes; top-down processes compensated for inefficient bottom-up processing due to the impairment of information at lower levels.

In another experiment (Klein, 1976) the secondary task was a digit recall or rehearsal task which subjects performed on both random and coherent WBT passages. The principle reason for including the memory task was to examine the effects of context when attention was distracted from the task of word identification. The results showed that increased attentional demands at the orthographic level reduced the efficiency of semantic analysis, however, an effect for context utilization was still found.

Schwartz (1977) viewed the results of Klein's experiments (1976), and the findings of his own research, as empirical support for the theory of "independent knowledge sources" (Rumelhart, 1977). Regarding those
experiments, Schwartz stated that "The rehearsal task reduces top-down processing while leaving other codes intact; the orthographic manipulation reduces the effectiveness of bottom-up processes, but not syntactic/semantic codes. Both results confirm the independence of knowledge sources" (p. 18). In addition, Klein (1976) pointed out that Posner's and Snyder's (1975) conceptualizations, including their notions of operational memory and compensatory processing, offer further explanatory value for his WBT findings, and for WBT performance in general. Klein's comments are these:

In the word boundary paradigm, the detection of semantic and syntactic rules and regularities is the basis of the use of contextual information... Subjects use these regularities to make inferences and predictions that improve performance in coherent conditions. The more words that subjects can maintain in operational memory, the greater the likelihood that such rules and regularities will be discovered. The more regularities that the subject holds in operational memory, the more accurate are his predictions... when a subject tentatively identifies a word, he verifies this by checking out the next few words; the more words he can hold in operational memory, the more effective his verification procedure. (p. 30)

Although theories of reading offer explanations for the effects of context on WBT performance, it must be noted that results obtained with this instrument should not be interpreted directly as an indication of the subject's behaviour in the actual reading situation. The WBT is a problem-solving task that elicits the performer's knowledge, skills, and strategies that may be used by individuals during reading, but not necessarily, and perhaps not to the same extent. As Stanovich (1980) cautions, although an individual may display knowledge and use of contextual constraints when called upon to do so, this does not imply that the same individual uses or relies upon such knowledge when reading. To clarify
this issue more research is needed to understand the relationship between reading comprehension and context utilization as defined operationally in terms of word-boundary task performance. A step in this direction is an investigation of the validity of the WBT as a measure of that construct.

Measurement of Context Utilization

Research employing several different versions of the WBT has demonstrated the facilitative effects of context utilization on word identification. In one of these studies (Klein, Klein, and Bertino, 1974) a context utilization measure was obtained to investigate the relationship between that factor and teacher ratings of subjects' reading competence. These researchers proposed that "the comparison between R [random] and C [coherent] passages reflects the extent to which subjects use the contextual information in the coherent passages. This can be measured simply in terms of the increment from the R to the C passages, using the R as a baseline" (p. 80). This same rationale was followed by Schwartz (1977) to generate a similar measure reflecting context utilization.

The potential value of the difference score for measuring context utilization in the manner prescribed here, must be examined critically in terms of the limitations of the WBT. Traditionally, difference scores pose theoretical as well as practical problems. Technically, the original scores are more reliable than the difference between them; these practical limitations can be overcome through the employment of

---

2To investigate this relationship Klein et al. followed appropriate statistical procedures for eliminating problems caused by difference scores. The analysis showed a significant positive relationship between context utilization and teacher ratings of subjects' reading competence (r = .47, p < .01).
appropriate statistical techniques. Theoretically, however, the difference score is more worrisome. While something is assumed to be gained it is not always possible to define what that something is. The following discussion approaches the question of measuring context utilization from both theoretical and methodological viewpoints. The issues addressed were taken into consideration along with the findings of this research to determine the construct validity of the instrument.

Limitations of a Difference Score: The WBT as used in this study, comprised one coherent passage and a randomized version of this passage. Each subject responded to both passages, providing two WBT measures. These repeated measures are straightforward enough in and of themselves. It is reasonable to posit that performance on the random passage reflects one's ability to identify words, given only orthographic constraints, imposed by featural, letter-level, letter-cluster, and lexical-level knowledge sources. A score on the coherent passage is measuring word identification within a context; context in this situation must be defined as the presence and overlap of orthographic, syntactic and semantic constraints. Both passages share orthographic constraints in common. To a certain degree, even in the random passage, semantics interacts with orthographies since the letter sequences do form words. There are, however, no meaningful links between or across words, and therefore, sequential constraints that constitute a coherent text are lacking. It is the meaningful ordering or structuring of words that provides context as defined previously; but, it is the individual who must perceive or construct the context.
In comparing performance on the random and coherent passages, the difference can be attributed to the effects of constraints beyond those found to be mutual to both passages (see Figure 2.1). In the random passage, subjects can rely only on orthographic constraints (white area) to identify words, while in the coherent version all three cueing systems are available to the individual for performing the same task (striped area). If subjects do in fact make use of all of these constraints, they have at their disposal information in the coherent passage that is absent in the random version of the passage (shaded area). The difference in performance on the two versions, however, partials out the effects of orthographic constraints and the remaining information is attributed to the overlap of syntactic and semantic constraints (checked area).

Figure 2.1
The Overlap of Linguistic Constraints
While this formulation of the theoretical significance of the difference score is tenable, there are alternative hypotheses. In the coherent passage, individuals may not be interacting with all three systems to the same degree. More information in fact may be coming from one source of information or another. One could argue, for example, that an individual making use of knowledge of syntactic and orthographic constraints alone would have a difference score that reflected the use of syntactic constraints only. Indeed, the meaning of a difference score may vary with different individuals; disparities among individuals may be not only a function of factors involving context utilization, but also of the sheer idiosyncratic responses of test-takers.

Supposing that the difference score is a reflection of individuals' use of context, its utility must, nevertheless, be demonstrated. Any response measure can be understood and employed effectively, only if valid assumptions concerning the scale and dimensions of measurement are established. These criteria are required to evaluate the distribution and magnitude of the scores—to judge, for example, whether a difference score of 20 points indicates a "non-chance" gain due to context utilization, and if so, whether this measure departs significantly from an increment of 10 points, and by how much. Such issues regarding both the nature and magnitude of the difference score, must be addressed.

The significance of a difference score would also be highly dependent upon the methodology employed with the WRT. This point is reinforced by considering for example, the rules researchers must follow strictly, to construct a valid passage for cloze task performance (see Bormuth, 1968). It is therefore necessary to posit certain a priori assumptions
regarding the potential effects of the variables involved, whenever the WBT is employed. In this study, hypotheses concerning the influence of the experimental variables were both theoretically and empirically based. It was hypothesized, for example, that performance on the coherent passage would be significantly better than performance on the random, regardless of treatment condition. Because of the intention to explore the difference between random and coherent performance as a measure of context utilization, this was the variable of main interest. In developing some notion of the potential significance of this difference score, performance on the random and coherent versions of the passage were first viewed separately with respect to the individual performance conditions.

Regarding the random version (see Figure 2.2), there was no expectation that performance would vary due to contextual clueing or boundary markers. For most intents and purposes, the random versions were considered a baseline for their coherent counterparts. In terms of performance

<table>
<thead>
<tr>
<th>WBT Performance Conditions</th>
<th>No Clue</th>
<th>Clue</th>
<th>No Clue</th>
<th>Clue</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Markers</td>
<td>No Markers</td>
<td>Markers</td>
<td>Markers</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2.2
Random Version - Performance Conditions

on the coherent version, however, there appeared to be potential for progression along a continuum (see Figure 2.3). Placed at the one pole were the conditions with no contextual clueing and no sentential boundary markers, and at the other extreme, conditions in which both of these
features were present. Apparently, there was potential for movement in a positive direction from the least constrained and lowest approximation to reading, to the most constrained and greatest approximation to reading. No hypotheses were made concerning the other variable combinations, except that they appeared to belong somewhere in the middle.

![Table and Diagram]

Figure 2.3
Coherent Version - Performance Conditions

Next, both versions of the passage were considered in terms of their influence upon each other. More specifically, should performance increase whenever either version of the passage was presented last, a practice effect would be implied. A pattern such as this would be detrimental to measuring context utilization in the prescribed manner. It would, indeed, confound the significance of the difference score, rendering an operational definition of context utilization meaningless. On the other hand, if significant gains for coherent over random passage performance were maintained throughout all conditions, the parallel view of random and coherent performance might be extended to a continuum (see Figure 2.4). If progression in a positive direction from performance on
the random passage to that on the coherent were established, reliable conclusions could then be reached concerning the potential use of the WBT as a measure of context utilization. Further statements concerning the instrument's construct validity, however, would still be dependent upon the nature and strength of the relationships between context utilization and other theoretically related factors.

![Projected Performance Continuum](image)

Despite the limitations outlined here, it is still possible that the difference score reflects context utilization as measured by a WBT. Moreover, any conclusions reached by this research cannot nullify or confirm the tenability of that hypothesis.

Another measure to be considered, not as an alternative to, but in conjunction with the difference score, is an index of differential speed. This measure (discussed briefly in the following paragraphs) would be subject to the same scrutiny as the difference score, but might afford some additional information concerning WBT performance and context utilization.
**Differential Speed Ratio:** The WBT is a speed test. The rationale behind imposing time constraints is that, given sufficient time, the task would be so trivial that nearly all subjects would receive perfect scores. If no reliable dispersion of scores could be obtained, the WBT would have no properties as a measure of context utilization. In exchange for the discriminatory and predictive power gained through time constraints, however, the accuracy of responses may be lost. Although accuracy is emphasized over speed in WBT instructions, the extent to which individual subjects adhere to those rules, by making every decision their most accurate, is not known. Speed, therefore, is a variable upon which performance is contingent.

It is impossible to determine the effects of speed without varying the time element. Doing so, however, for the purposes of this study, meant altering also the structure of the factor of interest; namely, context utilization. An indication of an individual's use of context that would take into account the speed involved, however, would be a ratio of the number of correctly identified words in the coherent passage to the same tally in the randomized version. This measure would also provide additional information not reflected in the difference score. More specifically, two individuals could have the same difference score of 10 points, for example, yet differ in terms of their internal speed—20/10 as opposed to 80/70. This index of differential speed would provide information about the original scores that is lost in the difference score. These measures together would define context utilization, provided the outcomes of the WBT experiment confirmed the expectations outlined above.
Summary

This section reviewed previous investigations with the WBT that lend empirical support for the methodological artifacts questioned in this research. Those studies, as well as the theories discussed in the first section of this chapter raise questions about the word-boundary task's validity as a measure of context utilization. Finally, the theoretical and practical limitations of the instrument as it was used in this research were also discussed.
CHAPTER III
PROCEDURES

Population and Sample

The target population consisted of students enrolled in intermediate courses of Italian at the university level. In order to qualify for the intermediate level, as the term is used here, the student must already have completed at least one introductory course in Italian. This level was chosen for both theoretical and practical reasons. Many students who enroll in language courses do so in order to fulfill certain graduation requirements. Only a small percentage pursue the study of Italian beyond this stage. Therefore, for the majority of students, the completion of the second course marks a half-way point in their exposure time to the language. The intermediate level was also selected because of the increase in the attrition rate after each course. Unfortunately, all students who begin a language do not schedule each sequence according to successive academic terms. There are usually fewer sections offered each term for levels beyond the first course. Therefore, given that exposure to the language is brief for the most part and that the availability of subjects is limited, the intermediate level seemed most appropriate for the purposes of this study.

The sample consisted of 114 intermediate students from three universities. The Ohio State University, Miami University, Oxford, Ohio and The State University of New York at Buffalo, formally consented to
participate in the research. At each university all students enrolled in classes of the specified level during the Winter and Spring terms of 1981 were requested to participate. The convenience sample described was believed to be representative of the previously mentioned target population.

Some differences among the participating universities relative to the actual time spent in the classroom, average class size, and the content covered in the textbooks used at the universities were expected.

[For a breakdown of other descriptive variables by university see Table 1, Appendix D.] Two of the universities in the sample offered semester hours (14 weeks) while the other followed a quarter schedule of 10 weeks duration. Therefore, the number and length of class sessions for those students included in the sample varied from 5 days per week for 50-minute periods each to 2 days per week, 180 minutes each. In addition to these differences, class size ranged from 6 to 20 students in various classes.

[The basic textbook used in individual classes also differed among and within the universities. The two most widely used texts were Basic Italian, Fourth Edition, by Speroni and Golino; New York, Holt, Rinehart and Winston, 1977, and Oggi in Italia (Merlonghi, Merlonghi, and Tursi; Boston, Houghton Mifflin Company, 1978).] The majority of classes utilizing Basic Italian were scheduled to cover the first 24 chapters of the text prior to data collection and those using Oggi in Italia were to have completed 20 chapters in that text. In the development of the language proficiency tests both of these textbooks were taken into account; their contents are described in detail within the section on Instrumentation.
Instrumentation

All instruments were either specifically developed or adapted by the researcher to measure the constructs of interest to this study. In an effort to construct reliable instruments the guidelines advocated by many authorities in psychometric research, testing, and evaluation were adhered to carefully. All instruments were pilot tested at the Ohio State University during the Summer Quarter 1980 and the Fall Quarter of 1980. The following sections describe in detail the steps taken in the development of each instrument.

Reading Comprehension Tests

Reading comprehension was measured in terms of both the free recall and recognition of two reading passages. Initially, three passages were written based on articles from Italian magazines and adapted to the proficiency level of the subjects under study. The language and writing style of the researcher was authenticated by a native speaker of Italian. The approximate length of each passage was 250 words.

A multiple-choice recognition test was constructed in English consisting of 5 items for each passage. For each item, one correct answer and three distractors were provided. In order to determine the plausibility as well as the equal probability of all alternatives being chosen, the 15 items were administered to a group of 17 people who had not read any of the passages. Based on these data, revisions were made. Each passage was then pilot tested in classes of Italian to evaluate (i) the appropriateness of the passages, (ii) the length of time necessary to read and respond to the tests based on each passage, (iii) the
reliability of the recognition tests, and (iv) inter-rater reliability for scoring the recall protocols.

Subsequent to each reading, subjects were asked to write a summary in English of what they had read and to rate the passage as to its difficulty on a 5-point scale. Immediately following, a five item multiple-choice test based on the passage was administered. The ratings along with the actual feedback from the recall and recognition tests assisted in the selection of two reading passages for use in the main study (see Appendix A).

The mean difficulty ratings for the two acceptable passages were 3.59 and 3.51 respectively. With respect to the recognition tests, one point was granted for each correct answer, allowing a range in scores from zero to five for each passage. The recognition test for Passage 1 had a mean of 1.51 and a standard deviation of 1.14. For Passage 2, the mean was 2.17 with a standard deviation of 1.08. Reliabilities for each of the recognition tests and for the total ten items combined were obtained using Kuder-Richardson 20, a version of Cronbach's Alpha appropriate for dichotomous items. The sub-program RELIABILITY in SPSS (Statistical Package for the Social Sciences, 1980) computed reliabilities of .45 and .46 for the tests derived from each of the two readings, and .60 for the 10 items together. Based on these findings, and additional information such as item intercorrelations, five of the items were revised (three from the first passage, and two from the second passage) for use in the main study.

The propositions or idea units contained in the individual passages were identified prior to scoring the recall summaries (see Appendix A).
Two scorers read each summary independently for the pilot test data and followed the list of propositions for scoring the summaries. The scorers gave one point for each correctly recalled main idea. To determine the inter-rater reliability, Pearson Product-Moment correlations between the two sets of scores were calculated. A correlation of .97 was considered strong enough to warrant only one scorer for the main study. As a further check, however, 20% of the recall protocols in the main study were scored by another rater and reliability was once again assessed.

The procedures just described for scoring the recall and recognition tests were followed in the main study. The final recall and recognition scores, however, were obtained reflecting performance on both passages. To ensure equal weighting of each passage, data transformations were made in the following manner. Initially, each subject received four raw scores (2 measures of recall and 2 measures of recognition). Each of the four raw scores were then converted to standard scores. An average of the standard scores were obtained for recall and recognition respectively. The final average was once again transformed to a standardized score. Each subject therefore had one recall and one recognition score.

**Cloze Test**

The cloze procedure was employed to measure an individual's ability to infer meaning from text written in a second language. In constructing the test, an Italian prose passage, approximately 100 words in length, was prepared. To serve as an introduction, the first sentence was left intact. Thereafter, every seventh word was deleted producing a total of 14 blanks. This deletion rate was chosen with the time constraints, difficulty index, and discriminatory power of the test in mind. Although
no evidence has been found favouring a particular deletion pattern appropriate for use with non-native students, research with native speakers of English has most often reported every fifth-word deletions. This practice has been based to a great extent on MacGinitie's assertion that the ability to restore context is not facilitated by deletion rates beyond five words (1961, p. 127). Taylor (1953) however, who introduced the cloze procedure, compared various deletion rates and expressed a preference for deleting either one in five or one in seven words. In a study of cloze test performance with bilinguals, Carroll (1959) deleted every tenth word based upon the belief that "a lower rate of deletion might be more effective for measuring individual differences in language skill" (p. 21). Because text written in unfamiliar language produces some degree of ambiguity in itself, it was felt that a distance between blanks, greater than five words, should be maintained. Deleting every seventh word, in the passage developed for this study, allowed various parts of speech to be sampled and placed realistic demands on the subjects in the time segment available for testing.

The cloze passage is presented in Appendix B. Subjects were instructed to fill in each blank with an appropriate English and/or Italian word that best completed the context of the paragraph. Ten minutes were allotted for the test. Responses were scored according to an "acceptable" scoring method. More specifically, any word synonymous with the original or that made sense grammatically and semantically in the context of the passage, was counted as correct. Missing data were scored as zero. Oller (1979) reported that, with non-native speakers, this scoring method was superior to the exact word scoring procedure.
The cloze test reliability for pilot study data using Kuder-Richardson 20 was .75. The procedures just described were followed for collection of data in the main study.

Tests of Language Proficiency

Proficiency in Italian at the intermediate level was assessed in terms of performance on tests of syntax and vocabulary. Both tests were developed by the researcher following a multiple-choice format. For these instruments, items were selected that reflect the actual content covered in Basic Italian and Oggi in Italia prior to the time of testing. The following section is offered as a general description of the similarities and differences between the two texts for those who are unfamiliar with their contents and formats. It also presents the grammar points shared by both texts within the chapters of interest, as well as the percentages of words for the various parts of speech reflected in the vocabulary.

Textbooks

Both Basic Italian and Oggi in Italia present the basic vocabulary and grammatical constructions characteristic of introductory second-language textbooks. Their formats, however, are somewhat different.

Basic Italian consists of one introductory lesson on pronunciation, 36 regular chapters, and 9 review sections. Each chapter begins with a reading or dialogue followed by vocabulary and idiomatic expressions, grammatical explanations, pertinent exercises and a section of questions designed to stimulate conversation. After each sequence of 4 chapters, a "Ripetizione" appears that provides exercises for review of the preceding chapters. Following this section is a cultural reading.
Oggi in Italia contains 2 preliminary lessons (useful expressions) and 28 regular lessons. Each lesson begins with a dialogue or reading followed by a cultural note, vocabulary, substitution drills, guides for oral or written work, vocabulary expansion exercises, grammar explanations and related exercises. In addition, the first 17 lessons provide a section on pronunciation and even-numbered lessons include a review of basic structures and vocabulary.

For the purposes of this study, the first 24 chapters of Basic Italian were reviewed as well as the first 20 chapters of Oggi in Italia. Table 3.1 lists the mutual grammar points of the two textbooks and the chapters in which they appear. With regard to the vocabulary presented in these chapters, Table 3.2 gives a breakdown of the percentages for the various parts of speech, the projected number of words from each group for both 75 and 40 item tests, and the actual frequencies reflected in the final 40 item vocabulary test.
### TABLE 3.1
SUMMARY DESCRIPTION OF TEXTBOOK MATERIAL

<table>
<thead>
<tr>
<th>GRAMMAR POINT</th>
<th>CHAPTER:</th>
<th>BASIC ITALIAN</th>
<th>OGGI IN ITALIA</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOUNS:</td>
<td>number/gender</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>ARTICLES:</td>
<td>definite/indefinite</td>
<td>1,4</td>
<td>1,3,4</td>
</tr>
<tr>
<td>ADJECTIVES:</td>
<td>forms/position</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>possessive</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>demonstrative</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>PRONOUNS:</td>
<td>subject</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>possessive</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>reflexive</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>direct object</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>indirect object</td>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>demonstrative</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>disjunctive</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>VERBS:</td>
<td>present tense</td>
<td>2,3</td>
<td>3,5,9</td>
</tr>
<tr>
<td></td>
<td>future</td>
<td>6,7</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>present perfect</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>reflexive</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>imperfect</td>
<td>16</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>imperative (informal)</td>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td>POSSESSION:</td>
<td>di</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>INTERROGATIVES:</td>
<td></td>
<td>2,8</td>
<td>6</td>
</tr>
<tr>
<td>NEGATION:</td>
<td></td>
<td>3,10</td>
<td>11</td>
</tr>
<tr>
<td>PREPOSITIONS:</td>
<td></td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>ADVERBS:</td>
<td>mente</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>PARTITIVE:</td>
<td>ne</td>
<td>22</td>
<td>20</td>
</tr>
</tbody>
</table>
### Table 3.2

**Vocabulary Breakdown**

<table>
<thead>
<tr>
<th>Parts of Speech</th>
<th>%</th>
<th>75 Words</th>
<th>40 Words</th>
<th>Final Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>44</td>
<td>33</td>
<td>17</td>
<td>16</td>
</tr>
<tr>
<td>Verbs</td>
<td>24</td>
<td>18</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Adjectives</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Adverbs</td>
<td>10</td>
<td>7</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Prepositions</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Conjunctions</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Pronouns</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Articles</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

**Syntax Test**

The test for proficiency in Italian syntax was based on the grammar points presented to the students during the course of their study of the second language. The grammar which both texts have in common within the target chapters has already been presented in the section on Textbooks. These grammar elements were used in generating items for the syntax test and were tested within the context of an Italian sentence, in which relatively simple vocabulary was chosen to minimize semantic involvement. Seventy items were constructed in a multiple-choice format with four alternatives written in Italian. The syntax test was piloted to evaluate the reliability and difficulty level of items as well as the quality of the individual distractors. Kuder-Richardson 20 reliabilities were calculated with 28 cases. The results are presented in Table 3.3.
final 40 item test was used in the main study to assess proficiency in Italian syntax (see Appendix B).

### TABLE 3.3
SYNTAX TEST RELIABILITIES BASED UPON PILOT STUDY DATA

<table>
<thead>
<tr>
<th>Reliability Check</th>
<th>Item</th>
<th>Item-Total r range</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>70</td>
<td>.01 - .72</td>
<td>.924</td>
</tr>
<tr>
<td>2</td>
<td>61</td>
<td>.02 - .69</td>
<td>.911</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>.06 - .67</td>
<td>.909</td>
</tr>
<tr>
<td>4</td>
<td>45</td>
<td>.05 - .69</td>
<td>.887</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>.16 - .69</td>
<td>.892</td>
</tr>
</tbody>
</table>

**Vocabulary Test**

The vocabulary test comprised words drawn directly from the target chapters of the two textbooks of interest. Only those words found in both texts were considered. To ensure a representative sample of words from the vocabulary actually presented, frequency counts and percentages were obtained for the various parts of speech introduced in the texts. Words were selected randomly to produce a pool of 75 words reflective of the relative percentages of each part of speech. Each Italian word in the vocabulary test was followed by 4 English words (3 distractors and a key word representing one meaning of the target word). Pilot tests were conducted to determine the (1) reliability of the vocabulary test,
(ii) level of difficulty of the instrument, (iii) power or value of each distractor, and (iv) appropriate time segment necessary for the vocabulary test.

Three pilot tests were conducted in different groups. Reliabilities were assessed using the Kuder-Richardson 20 Formula. The results of these analyses are presented in Table 3.4. The reliabilities remained relatively stable. Selection of items for the final word pool was based upon the item-total correlations (see Appendix B).

**TABLE 3.4**

**VOCABULARY TEST RELIABILITIES BASED UPON PILOT STUDY DATA**

<table>
<thead>
<tr>
<th>Reliability Check</th>
<th>Sample n</th>
<th>Word Pool</th>
<th>Deletions</th>
<th>Word Count</th>
<th>Item-Total r range</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>75</td>
<td>34</td>
<td>41</td>
<td>(.36)-(.78)</td>
<td>.858</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>56</td>
<td>14</td>
<td>42</td>
<td>(.19)-(.69)</td>
<td>.841</td>
</tr>
<tr>
<td>3</td>
<td>29</td>
<td>47</td>
<td>7</td>
<td>40</td>
<td>(.02)-(.75)</td>
<td>.862</td>
</tr>
</tbody>
</table>

_**Word Boundary Task**_

The word-boundary task (WBT) was used in this study with the intention of (i) empirically investigating its operation in a second-language setting, and (ii) ultimately obtaining meaningful measures of context utilization. The following sections present a description of the materials, procedures and scoring methods associated with the WBT as it was used in the experimental phase of this study.
**Materials**

The materials consisted of two versions of the same passage, one coherent and the other its randomized version in which words were arranged randomly. Each line began and ended with a complete word, typed with equal spacing between letters and across words, lengthwise on standard size paper (22 x 28 cm). Capitalization was omitted.

There were 4 basic forms of the WBT each identical in format, that is, each contained a random and a coherent version of the passage. The forms differed in one or more of the following ways: the presence (1) or absence (2) of context clueing, and the presence (1) or absence (2) of boundary markers. The forms were color-coded yielding green, yellow, blue and pink forms. Each of these forms had one alternate form that was distinguished from the basic form only with respect to a third factor, presentation sequence: (1) random-coherent, (2) coherent-random. Controlling for this variable yielded a total of 8 forms that together exhausted all possible combinations of the levels of the three variables.

Table 3.5 describes the levels of each of the three variables of interest that are reflected in the individual forms. These 3-digit numbers served as codes for the 8 different forms. The acronyms used in describing the 4 context clue and boundary marker conditions are also presented in this table.
TABLE 3.5
DESCRIPTION OF WORD-BOUNDARY INSTRUMENT DESIGN

<table>
<thead>
<tr>
<th>WBT Form</th>
<th>Context Clueing</th>
<th>Boundary Markers</th>
<th>Presentation Sequence</th>
<th>Condition Acronym</th>
</tr>
</thead>
<tbody>
<tr>
<td>GREEN</td>
<td>111</td>
<td>1</td>
<td>1</td>
<td>+Clue+Markers</td>
</tr>
<tr>
<td></td>
<td>112</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>YELLOW</td>
<td>121</td>
<td>1</td>
<td>2</td>
<td>+Clue-Markers</td>
</tr>
<tr>
<td></td>
<td>122</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>BLUE</td>
<td>211</td>
<td>2</td>
<td>1</td>
<td>-Clue+Markers</td>
</tr>
<tr>
<td></td>
<td>212</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>PINK</td>
<td>221</td>
<td>2</td>
<td>2</td>
<td>-Clue-Markers</td>
</tr>
<tr>
<td></td>
<td>222</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Factor Codes: 1 = yes 1 = present 1 = random-coherent
2 = no 2 = absent 2 = coherent-random

Eight booklets were prepared corresponding to the eight distinct forms. Each booklet, regardless of condition, contained a page of overall instructions and practice exercises identical across forms. Following this page, a pre-task instruction was presented. This served as (i) a clue to context in the appropriate conditions, and/or (ii) an explanation of the presence of two consecutive hyphens (− −). More specifically, if (− −) was encountered in the task, subjects were told to consider this either a "free slash" (random version) or a break between sentences (coherent version). The overall instructions are presented in Appendix C.
along with the pre-task instructions and the versions of the passage with and without boundary markers. The latter are form specific and can be identified by their color codes.

**Procedures**

Test booklets containing one of the 8 forms of the WBT were distributed randomly to subjects in each class at all universities. As explained in the overall instructions, the task for both passages across all forms was identical. Subjects were instructed to place slashes between word boundaries as quickly as possible, emphasizing accuracy over speed. The process for cancelling an incorrectly-placed slash was described and practice exercises in English and Italian were provided. Subjects were informed that no extra letters or nonsense words had been built in. They were told to work from top to bottom, left to right.

Finally, to introduce the term "free slash", subjects were told that words did not carry over to the next line and it was therefore unnecessary to slash the last word in a line (see Overall Instructions, Appendix C).

Questions regarding the procedure were permitted only during the overall instructional period. After this step, subjects read silently their pre-task instruction (approximately 30 seconds). A signal was given to start and stop each task. Approximately 45 seconds were allowed between passages to read the intermittent pre-task instruction introducing the next passage. Equal time (3 minutes) was allotted for proceeding through each passage.
Scoring

Both the random and coherent versions of the passage were scored for the number of correctly identified words. For the pilot test data the random passage was scored by (i) accepting any Italian word as correct and (ii) granting points only for the same words that appear in the coherent passage. The correlation between these sets of scores from the two scoring methods was calculated using Pearson Product-Moment procedures. The highly significant correlation (.97) indicated the compatibility of the two methods and justified using the more simple "same word" process of scoring for the main study.

To equalize the scoring procedures across all forms of the WBT, boundary markers (there are eight that appear in the Green and Blue forms) were counted as existing in the other forms as well. In terms of scoring, therefore, the same number of words to be identified existed for all passages on all forms. This scoring method did not partial out the potential effects of the boundary markers, because that condition was operative at the time of testing. If this factor were facilitative, its effect would be reflected in significantly higher scores for groups experiencing that treatment.

Data Collection Procedures

Data were collected at the universities by the researcher during the Winter and Spring terms (a period extending from March 5 to June 3, 1981). At each institution, the precise dates were scheduled during the final two weeks of the terms in which the selected level courses were offered. Data on all instruments were collected during two class sessions according to the following time schedule:
DAY 1

1. Reading Passage 1: reading: 10 minutes
   recall: 7 "
   recognition: 5 "

2. Syntax Test: 15 "

3. Cloze Passage:
   Total: 47 "

DAY 2

1. Word-Boundary Task: overall instructions: 7 minutes
   pre-task instructions: 2 "
   random version: 3 "
   coherent version: 3 "

2. Vocabulary Test: 10 "

3. Reading Passage 2: reading: 10 "
   recall: 7 "
   recognition: 5 "

   Total: 47 "

The sequence of tests within and between days remained fixed. This decision was made upon consideration of the following conditions:

(i) the different time allotments for the various tests along with the additional constrictions of class time-periods delimited the possibility of all combinations of test sequences. More specifically, if two reading passages had been given on the same day, a third day would have been necessary to complete the testing;

(ii) the different time allotments for the various tests necessitated the same testing order for all subjects within a class;
(iii) the number of classes that participated in the study was insufficient to control efficiently for potential order effects;

(iv) finally, in an effort to distribute the potential order effects, randomization of test orders across classrooms was considered. This option, however, was undesirable because it could have caused sequence effects to become confounded with classroom effects. Furthermore, no specific information concerning the effects of test sequence would have been afforded by this method.

The proposed order separated the two reading passages, the two tests of language proficiency and the two tests of reading strategies, providing some variety on each testing day. It was hoped that this spacing lessened the monotony and tedium of the testing sessions.

Prior to the days of testing, instructors of each class were asked to solicit the cooperation of their students by sharing with them a brief description of the study. At that time students were asked to sign a release form provided by the Bureau of Human Subjects, confirming their willingness to participate in the study.

Each day, test booklets containing the three tests ordered in the manner proposed were distributed to the subjects. Students were requested to write their name on a cover sheet and to supply some brief personal data (see Appendix D). The first test began immediately. Subjects were asked to read along with the researcher the instructions for each test and to start and stop each test when the signal was given. Test booklets were collected at the end of each class session.
Feedback from all tests was left to the discretion of the individual class instructor. After each data collection period, test copies and answer sheets were provided, that could be distributed among the students for use in class discussions and review. In addition, students were informed that they could contact the researcher for personal results. A summary of the results of the study was made available to the instructors subsequent to the data collection period.

Main Study Reliability Assessments

Reliabilities for each of the tests developed by the researcher were reassessed using the data from the 114 subjects who participated in the main study. The subprogram RELIABILITY in SPSS reported Kuder-Richardson 20 reliability coefficients of .87 for both syntax and vocabulary tests (40 items each), .85 for the 14-item cloze test, and .57 for the combined items of the two recognition tests. The reliability quoted for the recognition test clearly indicated a lack of internal consistency among the ten items. If individual items were deleted, however, K-R 20 as estimated by the computer program, would not have increased markedly. Given these circumstances, all ten items were retained for calculating the recognition score. Nevertheless, the somewhat low reliability of this instrument raised the question of the test's validity and served as a forewarning that statistical significance would be harder to achieve with the recognition measure.

All recall summaries for the main study were scored by the researcher following a list of main propositions for each passage. Twenty percent of the total number of summaries (228) were selected at random to be scored by the same rater who assisted in the pilot studies, and
agreement between raters was assessed once again. Pearson Product-Moment correlations between scores for each reading passage were .98.

Design and Statistical Analyses

The present study was both descriptive and experimental in nature. The descriptive aspect proceeded in two phases, the latter of which was contingent upon the findings of an intermediary experimental phase. This section discusses the design in terms of these three stages that correspond to the purposes of the study. All aspects of the study utilized data from the same subjects. Since these subjects were sampled from different universities, data on all measures were first tested for differences among separate sub-samples. In light of the results of those preliminary analyses adjustments were made in addressing each of the research questions. These modifications are alluded to in the following sections, but are described in greater detail in Chapter IV.

Descriptive Phase I: Data on five measures were obtained; namely, recall, recognition, the ability to infer, and proficiency in syntax and vocabulary. No physical manipulation of variables was exercised in obtaining these responses. The data from these criteria were used to explore Purpose 1 of the study. Specifically, the following question was addressed:

- What is the nature and strength of the relationship between measures of (i) language proficiency and (ii) the ability to infer meaning and measures of reading comprehension?

Multiple correlational procedures were employed to examine the joint contributions of language proficiency and the ability to infer meaning to reading comprehension. Measures of syntax and vocabulary skill and the ability to infer were considered independent variables, while the two
dependent measures were recall and recognition. Canonical Correlation was performed to determine the number of significant canonical vectors, the strength of the squared canonical correlations, and the amount of variance in the dependent variable set that could be explained by the shared variance in the independent variable set. Follow-up procedures to this analysis included multiple regression analyses, in which sample effects were isolated and the same independent variables were explored for their contributions to recall and recognition, viewed separately.

Experimental Phase: This aspect of the proposed research, corresponding to Purpose 2 of the study, was conducted to determine the extent to which context utilization could be measured meaningfully by means of a word-boundary task. It was designed specifically to address the following questions:

- Is performance on a word-boundary task affected by the inclusion of such variables as (i) verbal context clueing, (ii) sentential boundary markers, and (iii) the presentation sequencing of random and coherent versions of a passage?

- Are there differences in performance on a word-boundary task associated with differences in reading comprehension, language proficiency?

An experiment was designed in which subjects experienced one of the two levels of each treatment variable:

(A) Contextual Clueing
   1) yes
   2) no

(B) Boundary Markers
   1) present
   2) absent

(C) Task Sequence
   1) random-coherent
   2) coherent-random
In addition, all subjects were exposed to both levels of a trial factor (D) Task: 1) random, and 2) coherent. Each subject was assigned to two of the 16 cells of a Three Between-One Within-Subjects Design (Kennedy, 1979), shown below.

Figure 3.1
Three Between-One Within-Subjects Design Arrangement
The dependent variable was the actual score on the random and coherent versions of the passage. Data were subjected to 2 univariate analyses of variance with repeated measures, assigning subjects as the unit of analysis. The following sources of variation were tested in the initial analysis:

<table>
<thead>
<tr>
<th>Between Subjects</th>
<th>Within Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (Context Clue)</td>
<td>D (Task)</td>
</tr>
<tr>
<td>B (Boundary Markers)</td>
<td>AD</td>
</tr>
<tr>
<td>C (Task Sequence)</td>
<td>BD</td>
</tr>
<tr>
<td>AB</td>
<td>CD</td>
</tr>
<tr>
<td>AC</td>
<td>ABD</td>
</tr>
<tr>
<td>BC</td>
<td>ACD</td>
</tr>
<tr>
<td>S/ABC</td>
<td>BCD</td>
</tr>
<tr>
<td></td>
<td>SD/ABC</td>
</tr>
</tbody>
</table>

In addition, an analysis of variance was conducted, utilizing the same basic design, but building in a university factor to insulate against confounding sample effects.

**Descriptive Phase II**: This phase was concerned with Purpose 3 of the study and the following question:

- What is the nature and strength of the relationship between measures of (i) language proficiency (ii) the ability to infer meaning and (iii) context utilization and measures of reading comprehension?

The analytical procedures followed in addressing this question were based upon the outcomes of the experimental phase-data analyses regarding a measure of context utilization, and are discussed in the following chapter.
CHAPTER IV
RESULTS AND DISCUSSION

Preliminary Analyses

Data from 114 subjects were tested for university and teacher-class effects on seven criterion measures. Three groups ranging in size from 17 to 57 students comprised the university variable. The levels of the teacher-class variable were selected to distinguish among the classes of eight teachers who participated in the study. Data from the classes of two teachers who taught more than one section of the 102 course were pooled to maximize level n's for the teacher-class variable. Investigations of the effects of the university and class variables were handled separately because the number of classes as well as the number of teachers varied across universities.

Using the computer program MANOVA distributed by Clyde Computing Services (1973), two one-factor multivariate analyses of variance were performed to investigate differences among universities and among classes. The dependent variables were the scores obtained from the syntax, vocabulary, cloze, recall, and recognition tests, and the two measures of word-boundary task performance. As shown in Table 2 of Appendix D, the results of the MANOVA by university indicate significant effects on both of the roots possible in the analysis. Similarly, although 7 roots were possible in the analysis by class, only two roots were significant beyond the .05 alpha level. Multivariate analyses were
conducted to protect the alpha level; however, there was no substantive reason to pursue these results in a multivariate fashion. Therefore, follow-up procedures included univariate analyses of variance (see Table D.3) and post-hoc pairwise comparisons using Scheffé. Significant differences among universities and among classes were found on each of the criterion measures. One university sample and one class performed consistently better than one or more of the other groups on all of the language measures. Breakdowns of the descriptive statistics for each criterion measure by university and by class are presented in Tables D.4 and D.5.

As shown in Tables D.4 and D.5, strong distinctions among universities and classes on all performance measures are apparent. Although the effects of these assigned or classification variables are of little substantive interest, uncontrolled, they threatened an accurate assessment of the contributions of other variables under investigation. To minimize their influence and to control for confounding effects, university and class were therefore articulated in the design of the study whenever possible. The operation of these variables and the methodological modifications made to account for them are described in the following sections.

Word-Boundary Task Experiment

The experiment investigating the methodological artifacts associated with the word-boundary task was designed to control for one trial factor (performance on both random and coherent tasks), and three between-subjects variables (context clueing, boundary markers, and task sequence). To control the influence of sample differences, the
university variable was also incorporated into the design. The addition of this variable, however, increased the number of cells from 16 to 24, thereby drastically reducing the cell size from a minimum of 13 to only 2 subjects in some cells. Separate analyses were therefore conducted with and without the university variable. Both analyses of variance utilized a repeated-measures design (BALANOVA 5 in SOUPAC).²

All tables relevant to the original ANOVA excluding university are presented in the body of this chapter so that the reader may follow more easily the questions addressed specifically by the research. For the tables pertaining to the ANOVA including the university variable the reader is referred to Appendix D. With respect to the sources of variance common to the two designs, a comparison of the ANOVA summary tables confirms that the effects attaining significance were the same for both analyses.

University Effects in the WBT Experiment

The ANOVA summarized in Table E.6 reported a strong main effect for university \( F(2, 90) = 11.15, p < .0001 \) as well as the significant interaction of that variable with context clueing \( F(2, 90) = 3.91, p < .05 \). The means and standard deviations by university and context clue are presented in Table D.7 along with the overall level statistics.

Strictly speaking, the interaction effect merely indicates that the three university samples performed differently in the presence and absence of contextual clueing (see Figure 4.1). Another plausible

²When cell \( n \)'s are unequal and the design is unbalanced, BALANOVA 5 provides an unweighted means solution that yields an approximate analysis of variance. According to Kennedy (1973, pp. 445-446), the unweighted means approach is most appropriate when the inequality of cell \( n \)'s is due to circumstances unrelated to the nature of the treatment. This was the case in the present study.
University by Context Clue Interaction

Figure 4.1
interpretation of potential theoretical importance was entertained, but requires discussion of an observed attribute of the university variable.

Follow-up procedures (Scheffé test) to the significant university main effect revealed significant decreases in mean performance on the WBT in descending order from University I to III suggesting high, average, and low levels of foreign-language proficiency. The correspondence of the university variable to degrees of language proficiency was also supported by the results of the analyses conducted previously on the seven criterion measures. Without exception, University I exhibited performance superior to that of University III on all measures, suggesting high and low levels of proficiency respectively, while University II fell somewhere in the middle by performing equally well as University I on some measures and comparably to University III on others; thus, the interaction between university and context clueing may be interpreted as follows. In the "no clue" condition there was a steady decrease in performance from high to low levels of proficiency, while in the "clue" condition a significant drop from high to average was found, but the low group remained on a par with the middle group. As shown in Figure 4.1 (both clueing modes considered) the pattern of performance for Universities I and II (high and average proficiency levels) is relatively parallel, although I is scoring higher and the second group shows a greater drop in performance than the former in the "clue" condition. The presence of contextual clueing did not affect these two groups positively, whereas the reverse appears to be true for University III (the group that scored lowest on all criterion measures). This interaction was independent of specific task conditions, therefore, performance may be
accounted for on the basis of overall skill in word identification irrespective of random and coherent versions of the passage.

While contextual clueing enhanced performance for the lowest ranking group, it appeared to hinder the performance of the higher ones. One explanation for this outcome might be that, for the lower proficiency group, context clues provided much needed sources of information that facilitated that group's performance. On the other hand, the context statement produced, perhaps, little more than a distraction that slowed the performance of the high and average groups. The additional time spent in selection and sampling text may have caused delays in word identification that reduced their normal speed.

Fluency in the use of strategies could further explain the discrepancies found within and between the high and average groups in the two clueing modes. The average group's lack of proficiency or automaticity in using strategies may have accounted for their significant drop (11.2 points) in performance from the "no clue" to the "clue" condition. Proficiency in the use of strategies could also explain the 13.4 point spread in scores of the high and average groups in the "no clue" condition as opposed to a difference half that size (6.7 points) in the "clue" condition.

An interesting observation made during the testing sessions corroborates this supposition and advances yet another explanation for group differences. During the experiment it was noticed occasionally that subjects spent several seconds of the total 180 seconds allotted for each task actually scanning the text, as if reading it, before beginning to mark slashes. The color codes associated with the WBT forms readily
confirmed that those subjects were experiencing the "clue" condition present in the yellow and green forms. Although the group membership of those subjects was not systematically recorded, more frequent occurrences of this behaviour were observed at University II. Given a contextual framework, the higher-level groups may have responded to the task less in terms of a word-identification exercise, and more in terms of a reading activity, while the low proficiency group faithfully pursued the word-identification instructions. If this were the case, the higher level groups may have utilized some of the strategies already discussed, or may have looked for units of meaning, as opposed to individual words, in order to grasp the whole of the passage. As mentioned previously, strategies that facilitate the normal reading process should not be expected to enhance performance on the WBT, given its task and time constraints. In fact, the demands of the WBT (to locate individual words), by nature of its format, would be less disruptive perhaps for those accustomed to reading in a word-by-word fashion.

The university by context clue interaction, although related only incidentally to the purposes of this study, raises some interesting theoretical issues and methodological questions. On the one hand, the interpretation of the clue statement's facilitative effect on University III is consistent with theories that emphasize the influence of higher-level processes on lower-level ones (Rumelhart, 1977; Smith, 1971). On the other hand, the performance of the other two groups provides evidence in support of LaBerge and Samuels' theory (1974) that attention to higher-level codes detracts from low-level processing when the latter operations are non-automatic (which is intentionally the case in the word-boundary
paradigm). Moreover, the university by context clue interaction contradicts the results of Schwartz's experiments (1977) in which no main or interaction effects were reported for the context statement variable.

In explaining the pattern variations across universities that occurred as a function of contextual clueing, a few alternatives have been offered, all of which rely on the assumption that a language proficiency factor was implicit in the university variable. In light of the inferences made in this respect, the aforementioned arguments should be regarded with caution.

Turning attention once again to Table D.6, the remainder of the effects bearing statistical significance were devoid of university influence. These include a main effect for the task variable as well as the interaction of task with task sequence, and the context clue by boundary markers by task interaction. These same effects also emerged in the ANOVA excluding the university variable (see Table 4.1) and are discussed in the following sections.

**The Task Factor**

The strong main effect for task indicates that regardless of the WBT form, performance was higher on the coherent version of the passage than on its random counterpart. The mean number of correctly identified words was 52.04 for the coherent task as opposed to 41.03 for the random task (see Table 4.2). This effect suggests that context, defined as the presence of syntactic and semantic constraints, facilitated word identification. In light of both first- and second-order interactions inclusive of the task factor, however, direct interpretation of the main effect must await interpretation of the joint effects.
### TABLE 4.1

**SUMMARY OF THE REPEATED-MEASURE ANOVA BY CONTEXT CLUE, BOUNDARY MARKERS AND TASK SEQUENCE**

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Subjects</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A (Context Clue)</td>
<td>1</td>
<td>9.38</td>
<td>9.38</td>
<td>.02</td>
</tr>
<tr>
<td>B (Boundary Markers)</td>
<td>1</td>
<td>67.67</td>
<td>67.67</td>
<td>.13</td>
</tr>
<tr>
<td>C (Sequence)</td>
<td>1</td>
<td>168.24</td>
<td>168.24</td>
<td>.33</td>
</tr>
<tr>
<td>AB</td>
<td>1</td>
<td>57.32</td>
<td>57.32</td>
<td>.11</td>
</tr>
<tr>
<td>AC</td>
<td>1</td>
<td>623.41</td>
<td>623.41</td>
<td>1.21</td>
</tr>
<tr>
<td>BC</td>
<td>1</td>
<td>467.48</td>
<td>467.48</td>
<td>.91</td>
</tr>
<tr>
<td>ABC</td>
<td>1</td>
<td>455.14</td>
<td>455.14</td>
<td>.89</td>
</tr>
<tr>
<td>S/ABC</td>
<td>106</td>
<td>54423.53</td>
<td>513.43</td>
<td></td>
</tr>
<tr>
<td>Within Subjects</td>
<td>114</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D (Task)</td>
<td>1</td>
<td>6882.57</td>
<td>6882.57</td>
<td>73.53**</td>
</tr>
<tr>
<td>AD</td>
<td>1</td>
<td>103.61</td>
<td>103.61</td>
<td>1.11</td>
</tr>
<tr>
<td>BD</td>
<td>1</td>
<td>34.06</td>
<td>34.06</td>
<td>.36</td>
</tr>
<tr>
<td>CD</td>
<td>1</td>
<td>8491.20</td>
<td>8491.20</td>
<td>90.71**</td>
</tr>
<tr>
<td>ABD</td>
<td>1</td>
<td>275.77</td>
<td>275.77</td>
<td>2.95*</td>
</tr>
<tr>
<td>ACD</td>
<td>1</td>
<td>28.80</td>
<td>28.80</td>
<td>.31</td>
</tr>
<tr>
<td>BCD</td>
<td>1</td>
<td>32.56</td>
<td>32.56</td>
<td>.35</td>
</tr>
<tr>
<td>ABCD</td>
<td>1</td>
<td>43.72</td>
<td>43.72</td>
<td>.47</td>
</tr>
<tr>
<td>SD/ABC</td>
<td>106</td>
<td>9922.17</td>
<td>93.61</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>227</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* $p < .08$

** $p < .0001$
TABLE 4.2
MEANS AND STANDARD DEVIATIONS FOR WORD-BOUNDARY TASK PERFORMANCE BY SEQUENCE AND TASK

<table>
<thead>
<tr>
<th>SEQUENCE</th>
<th>TASK</th>
<th>OVER TASK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random</td>
<td>Random</td>
<td>34.06</td>
</tr>
<tr>
<td>Coherent</td>
<td>(SD)</td>
<td>(12.60)</td>
</tr>
<tr>
<td>Coherent</td>
<td>Random</td>
<td>48.00</td>
</tr>
<tr>
<td>(SD)</td>
<td></td>
<td>(16.37)</td>
</tr>
<tr>
<td>Over</td>
<td>Random</td>
<td>41.03</td>
</tr>
<tr>
<td>Sequence</td>
<td>(SD)</td>
<td>(16.15)</td>
</tr>
</tbody>
</table>
Task By Task Sequence Interaction

Figures 4.2 and 4.3 demonstrate the effect of task sequence on task performance. Both graphs depict the same interaction from different perspectives. While the first offers a clearer view of the separation in task performance and the second allows a better comparison of the patterns of performance for the two sequence conditions, the same conclusion can be drawn from either graph: the interaction of task with task sequence implied a practice effect. Performance on a given task was favoured when that particular task was presented last in the sequence. When the coherent task was in final position performance on that task rose 11 points. Similarly, performance increased 14 points on the random task when that task occurred last. Even given this increase for the random task, however, a performance level only comparable statistically to performance on the coherent task in first position was achieved. In other words, performance under coherent-random sequence conditions was relatively stable across tasks (the mean performance on the two tasks differed by only one point). By contrast, under random-coherent conditions, a significant separation of 23 points was found between random and coherent task means.

These findings show that performance on both tasks was enhanced as a result of varying the sequence of tasks, yet performance on the coherent task was not exceeded significantly by random task performance under either sequence condition, although the reverse was observed. This explains the absence of a main effect for sequence. Furthermore, when each passage was in first position in terms of sequence, the means for random and coherent tasks were 34.16 and 46.83 respectively and when both were
Random Task  
Coherent Task  

Figure 4.2  
Task By Task Sequence

Random-Coherent Sequence  
Coherent-Random Sequence

Figure 4.3  
Task Sequence By Task
in last position, the spread was 11.4 points. In both instances performance was superior on the coherent task, all but ruling out a practice effect.

The results would suggest that context (present in the coherent version of the passage) still operated beyond the effect of sequence to facilitate performance on the WBT. While the task by task sequence interaction obscures the effect of context utilization as proposed by this study, there remains a reasonably viable facilitative effect of context on word identification.

Interaction of Task with Context Clueing and Boundary Markers

Further empirical support for the theory that context facilitates WBT performance was derived from the context clue by boundary markers by task interaction. This effect is of particular theoretical interest because it is independent of sequence effects just discussed. It should be noted, however, that although this three-way interaction was significant at the .05 alpha level in the ANOVA including the university variable, it only approached significance in the analysis summarized in Table 4.1. Nevertheless, it was felt that the proximity of these results necessitated the discussion of this effect.

The interaction is represented visually in Figure 4.4, with all combinations of the context clue and boundary marker factors arranged within the graph. As indicated in the accompanying grid, these combinations of variables correspond directly to individual WBT forms. This representation shows that mean scores for each task cluster into two groups. It is apparent that performance on the coherent task was much higher than that on the random task for all forms. Table 4.3 confirms that the means for
Figure 4.4

Context Clue By Boundary Marker By Task Interaction
TABLE 4.3

DESCRIPTIVE STATISTICS OF WBT PERFORMANCE
BY CONTEXT CLUE, BOUNDARY MARKERS AND TASK

<table>
<thead>
<tr>
<th>Context Clue</th>
<th>Boundary Markers</th>
<th>Random (Mean, SD)</th>
<th>Coherent (Mean, SD)</th>
<th>Over Tasks (Mean, SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Present</td>
<td>40.23 (15.10)</td>
<td>51.15 (22.29)</td>
<td>45.69 (16.28)</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>40.89 (17.79)</td>
<td>54.68 (18.03)</td>
<td>47.79 (15.17)</td>
</tr>
<tr>
<td>No</td>
<td>Present</td>
<td>39.97 (11.95)</td>
<td>52.60 (21.21)</td>
<td>46.29 (14.50)</td>
</tr>
<tr>
<td></td>
<td>Absent</td>
<td>43.03 (19.64)</td>
<td>49.71 (19.28)</td>
<td>46.38 (17.80)</td>
</tr>
<tr>
<td>Levels:</td>
<td>Clue</td>
<td>40.56 (16.33)</td>
<td>52.92 (20.22)</td>
<td>46.74 (15.53)</td>
</tr>
<tr>
<td></td>
<td>No Clue</td>
<td>41.50 (16.12)</td>
<td>51.16 (20.16)</td>
<td>46.33 (16.06)</td>
</tr>
<tr>
<td></td>
<td>Markers</td>
<td>40.10 (13.51)</td>
<td>51.88 (21.59)</td>
<td>45.99 (15.28)</td>
</tr>
<tr>
<td></td>
<td>No Markers</td>
<td>41.96 (18.60)</td>
<td>52.20 (18.66)</td>
<td>47.08 (16.40)</td>
</tr>
<tr>
<td></td>
<td>Population</td>
<td>41.03 (16.15)</td>
<td>52.04 (20.12)</td>
<td>46.62 (15.78)</td>
</tr>
</tbody>
</table>
the coherent task are all located above the overall grand mean (46.62) while those for the random task all lie below the mean. Closer inspection of Figure 4.4 indicates that the points corresponding to means on the random and coherent tasks follow slightly different patterns. Those pertaining to the coherent task are relatively equidistant, while for the random task almost no differences are exhibited among "+Clue-Markers", "+Clue+Markers", and "-Clue+Markers" treatment means, but, the mean of the "-Clue-Markers" group stands apart from the other three. In addition, across tasks, the rank order of forms based upon average performance shifted as a function of clue and marker manipulations. This change in performance was greatest for the "-Clue-Markers" condition. Whereas performance on the random task was highest with no clue and no boundaries, under those same conditions performance on the coherent task was lowest. Using the "-Clue-Markers" condition as a point of departure for both tasks, performance of the other three treatment groups decreased for the random task, but increased for the coherent task.

In determining which differences in average performance were significant statistically, it was of interest to compare not only the means within each task, but also those means between tasks. Post-hoc procedures therefore included all possible pairwise comparisons. In selecting an appropriate multiple-range test, consideration was given to the large number of comparisons desired as well as the inequality of cell n's. The discrepancies among cell n's, however, were extremely slight; there were 28 subjects each in the "+Clue-Markers" and "-Clue-Markers" conditions, and 29 subjects in the remaining two treatment conditions. (These sizes were the same across tasks, of course, since task was a repeated
measure). The balance was such that for twelve of the twenty-eight comparisons (approximately 43%), equal-n conditions obtained. Under these circumstances, Tukey's HSD could be employed conservatively by entering the smallest n (i.e. 28 as opposed to 29) into the following equation:

\[ q_a = \sqrt{\frac{MS\ error}{n}} \]

With a total of 8 means and 106 error degrees of freedom, a critical value of 8.02 was calculated, a conservative estimate of the magnitude in mean differences required to maintain a .05 alpha level for the total number of comparisons possible in the design.

Using the criterion of 8.02, none of the differences between means within either task were significant. Therefore, independent of task, the context clue and boundary marker factors (as the absence of their main effects would verify) had no significant effects on performance. Also non-significant was the comparison between random and coherent mean scores for the no clue, no boundaries condition. All of the remaining differences between forms across tasks exceeded the critical value for significance. Thus, the means of "+Clue-Markers", "+Clue+Markers", and "-Clue+Markers" conditions for the coherent task were each significantly higher than all of the means for the random task. Furthermore, although the "-Clue-Markers" treatment mean for the coherent task did not differ significantly from that of its random counterpart, it was significantly higher than the other three treatment group means for the random task.

These follow-up procedures served to clarify the nature of the interaction. As shown in Figure 4.5, the context clue by boundary marker
Figure 4.5

Context Clue By Boundary Markers By Task: Another View
by task interaction was ordinal. The treatment means for the coherent
task were uniform statistically, as were those associated with the
random task. Clearly, the task factor played the major role in this
interaction.

Admittedly, the magnitude of task means was not constant at all
levels of the clue and marker variables. The interaction effect itself
implies a significant departure from parallelism. As shown in Figure 4.5,
regardless of the context clue condition, mean performance on the random
task was lower in the presence of boundary markers, whereas for the co­
herent task, performance decreased with boundaries in the "clue" condi­
tion, but increased with boundaries when no clue to context was provided.
These variations in the patterns of performance across tasks suggest that
context, supplied in the form of a cognitive or visual clue was more mean­
ingful or congruent in the coherent task than in its randomized version.
But, the pattern changes just described were only discussed in terms of
select comparisons. More specifically, it is difficult to argue for dif­
ferential effects of the clue or marker factors on task, when all but
one of the treatment comparisons made between tasks were significant.

One demonstration of context clue and boundary marker influence,
however, deserves mention. As pointed out earlier, no significant varia­
tion was found between task means for the no clue or boundaries condition,
but the coherent task means for the conditions containing a clue, markers,
or both, were significantly higher than the random task mean for the
"-Clue-Markers" condition. It was concluded, therefore, that the added
constraints of context clue and/or boundary markers did have a facilita­
tive effect on coherent task performance that went beyond the effect of
contextual constraints in the coherent passage. This effect only held true, however, when the coherent task means were contrasted with the random task mean for the "-Clue-Markers" condition. For the most part, the simple effects of task prevailed, overshadowing the operation of the other two factors involved in the interaction.

**Context Utilization and the Word-Boundary Task**

The second purpose of the WBT experiment was to define operationally a measure of context utilization. In the analysis of WBT performance, task emerged as the most salient factor, supporting the assertion of previous researchers that context facilitates word identification in the WBT. Despite the task by task sequence effect, findings of this study generally support that claim. However, they cast doubt on the procedure for measuring context utilization by means of the WBT.

The most obvious obstacle to measuring context utilization suggested by this research was the effect of sequence on task. For the purpose of illustrating sequence effects, difference scores were generated by subtracting the random task score from the score for the coherent task, and frequency distributions of those scores were obtained. As expected, sign reversals were immediately visible as a function of task sequence: positive scores resulted with random-coherent sequence, and when the coherent task preceded the random, approximately 54% of the scores were negative values. The descriptive statistics presented in Table 4.4 clarify this observation. Whereas the overall average increment in performance for the coherent task over the random task was approximately 10 points, this mean difference would have changed markedly had sequence not been an influencing factor.
TABLE 4.4
MEANS AND STANDARD DEVIATIONS OF DIFFERENCE SCORES
BY WBT FORM

<table>
<thead>
<tr>
<th>WBT Condition</th>
<th>Task Sequence</th>
<th>Means</th>
<th>SDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>+Clue+Markers</td>
<td>R - C</td>
<td>22.23</td>
<td>20.61</td>
</tr>
<tr>
<td></td>
<td>C - R</td>
<td>-0.38</td>
<td>2.99</td>
</tr>
<tr>
<td>-Clue+Markers</td>
<td>R - C</td>
<td>24.27</td>
<td>14.69</td>
</tr>
<tr>
<td></td>
<td>C - R</td>
<td>1.00</td>
<td>14.48</td>
</tr>
<tr>
<td>+Clue-Markers</td>
<td>R - C</td>
<td>28.36</td>
<td>12.61</td>
</tr>
<tr>
<td></td>
<td>C - R</td>
<td>-0.79</td>
<td>11.73</td>
</tr>
<tr>
<td>-Clue-Markers</td>
<td>R - C</td>
<td>18.07</td>
<td>11.26</td>
</tr>
<tr>
<td></td>
<td>C - R</td>
<td>-4.71</td>
<td>10.22</td>
</tr>
<tr>
<td>Over Forms</td>
<td></td>
<td>10.83</td>
<td>18.32</td>
</tr>
</tbody>
</table>
To a great extent, task sequence limitations are methodological in origin. Sequence, however, is not the only factor obscuring the meaning of a difference score. To illustrate, four subjects in this study were given difference scores of 9, 9, 6, and 3 respectively. These data would imply that the first two subjects made use of context equally and also utilized context more than the other two subjects. A somewhat different interpretation follows, however, upon examination of these same subjects' original scores for the coherent and random tasks: #1 \((17 - 8)\); #2 \((63 - 54)\); #3 \((35 - 29)\); and #4 \((31 - 28)\). Lost in the difference score is the information concerning the level at which an individual performed to begin with. The first two subjects had the same difference scores even though subject #1 scored far below the mean for each task and subject #2 scored well above the task means. In addition, the second subject's score of 63 points for the coherent task was very close to the maximum number of words identified by any subject given the 3-minute time allotment for each task. It appears that subject #2 might simply have run out of time, whereas subject #1 may have experienced real difficulties in identifying words in or out of context. One might wonder whether the difference scores of these two individuals are really equivalent, and if, in fact, they are measuring the same thing. To some extent, the combination of the difference score with the differential speed ratio described in Chapter II might serve to clarify the problems of information loss.

In addition to determining the content of the measure captured by the difference score, the magnitude of that score must also be taken into account. A criterion must be established enabling one to argue with
assurance that the difference score represents a "real" increment in WBT performance. The results of this study raised questions about the interpretation of such increments. To illustrate this point, the latter three subjects in the example offered above each responded to the "no clue, no boundaries" condition. As reported, post-hoc comparisons revealed that the difference between the random and coherent task means for that form was no greater than chance expectations. In this particular case, given no significant differences in the original scores to start with, the generation of a "difference" score as a measure of context utilization or any other construct would be an unwarranted leap. Granted, the data presented here were purposely chosen to demonstrate the extreme situation in which the random and coherent tasks were not different statistically. Nevertheless, these data raise an important question about the WBT. Such methodological factors must be taken into consideration.

Another relevant aspect of this research problem, of less immediate concern, however, is the scale of measurement of the difference score. More specifically, in reference to the example, one might ask whether the distance between scores of 9 and 6 is equal to the distance between 6 and 3. The original data suggest that subjects #3 and #4 were more alike in terms of WBT performance than subjects #2 and #3. In light of these observations, it is dubious that equal intervals between difference scores can be assumed. At best, the measure approaches an ordinal scale. While the present study failed to demonstrate the validity of the word-boundary task as a measure of context utilization, the study did uncover problems in both the assumptions about the increments represented by difference scores and about the scale of measurement these scores imply.
Relationship Between Measures of Language Proficiency and Reading Comprehension

The descriptive aspect of this study addressed questions of the relationship between sets of language proficiency measures and reading comprehension measures obtained from students of a second language. In order to investigate both the nature and strength of the relationship between these sets, canonical variate analyses were performed. For those unfamiliar with the method of canonical correlation, a brief description of the process is given in Appendix E.

Analysis and Interpretation of Between-Sets Correlations

At the onset of this study, two related questions were proposed, namely:

What is the nature and strength of the relationship between measures of language proficiency including (i) vocabulary knowledge, (ii) syntax knowledge, and (iii) the ability to infer meaning, and measures of reading comprehension?

and

What is the nature and strength of the relationship between measures of language proficiency including (i) vocabulary knowledge, (ii) syntax knowledge, (iii) the ability to infer meaning, and (iv) context utilization, and measures of reading comprehension?

Because of the unresolved issues pertaining to the measurement of context utilization presented earlier, however, the latter research question was altered to include instead, the two measures of WBT performance, each of which reflects skill in word identification.

To address the question of between-sets relations, three canonical analyses were performed. In each analysis, two measures of reading comprehension (recall and recognition) comprised the criterion variable set. The number of variables in the predictor set, however, changed for each analysis. Initially, only the measures obtained from the syntax,
vocabulary and cloze tests were entered. Subsequently, a fourth member (the score for performance on the coherent passage of the WBT) was added to the set. Finally, in the third analysis, the predictor set was composed of all of the above-mentioned variables plus the measure of random task performance. Varying the number of variables in the predictor (left) set had no influence on the number of possible new canonical functions, since the criterion (right) set was always the smaller of the two.

The correlation matrix presented in Table 4.5 includes all predictor and criterion variables. The matrix has been subdivided to illustrate the smaller matrices operated upon in the last canonical analysis in which the fullest predictor set was utilized. From this same matrix, however, the between and within-set matrices specific to each execution of the canonical can be inferred.

A multivariate routine, CANON, authored by W.W. Cooley and R.R. Lohnes (Multivariate Data Analysis, 1971) computed all of the canonical correlations. The results of these analyses appear in Table 4.6. The findings revealed that for each analysis only the first canonical variate (root) was significant. In other words, the source of variance extracted by the first pair of composites was significant statistically, however, a significant amount of shared variance was not found with the second pair in any of the analyses.

In the initial analysis the maximum correlation between the left and right sets was .79, and the proportion of variance explained in the right set, given the left, was 48%—a substantial amount of variance shared by the two sets. Table 4.6 further indicates, however, that these
### TABLE 4.5

**INTERCORRELATIONS AMONG MEASURES OF LANGUAGE PROFICIENCY, WBT, MEASURES, AND MEASURES OF READING COMPREHENSION**

<table>
<thead>
<tr>
<th>Measures</th>
<th>Syntax</th>
<th>Vocabulary</th>
<th>Cloze</th>
<th>Random</th>
<th>Coherent</th>
<th>Recall</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td>—</td>
<td>.77</td>
<td>.71</td>
<td>.32</td>
<td>.62</td>
<td>.66</td>
<td>.55</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>—</td>
<td>—</td>
<td>.75</td>
<td>.27</td>
<td>.51</td>
<td>.68</td>
<td>.58</td>
</tr>
<tr>
<td>Cloze</td>
<td>—</td>
<td>—</td>
<td>.30</td>
<td>.49</td>
<td>—</td>
<td>.72</td>
<td>.57</td>
</tr>
<tr>
<td>Random</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>.51</td>
<td>—</td>
<td>.31</td>
<td>.31</td>
</tr>
<tr>
<td>Coherent</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

$R_{xx}$ - within predictor set correlations

$R_{yy}$ - within criterion set correlations

$R_{xy}$ - between-sets correlations

$R_{yx}$ - transpose of $R_{xy}$
<table>
<thead>
<tr>
<th>Predictor Set</th>
<th>Root</th>
<th>Canonical Correlation</th>
<th>Squared Canonical Correlation</th>
<th>$\chi^2$</th>
<th>df</th>
<th>p &lt;</th>
<th>Redundancy for Criterion Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Syntax</td>
<td>1</td>
<td>.79</td>
<td>.62</td>
<td>107.92</td>
<td>6</td>
<td>.0001</td>
<td>.48</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>.07</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td>0.00</td>
</tr>
<tr>
<td>Vocabulary</td>
<td></td>
<td></td>
<td></td>
<td>107.92</td>
<td>6</td>
<td>.0001</td>
<td>.48</td>
</tr>
<tr>
<td>Cloze</td>
<td>2</td>
<td>.07</td>
<td>.01</td>
<td>1.59</td>
<td>3</td>
<td>NS</td>
<td>.00</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>107.92</td>
<td>6</td>
<td>.0001</td>
<td>.48</td>
</tr>
<tr>
<td>Syntax</td>
<td>1</td>
<td>.79</td>
<td>.63</td>
<td>110.82</td>
<td>8</td>
<td>.0001</td>
<td>.49</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>2</td>
<td>.12</td>
<td>.01</td>
<td>1.59</td>
<td>3</td>
<td>NS</td>
<td>.00</td>
</tr>
<tr>
<td>Coherent</td>
<td></td>
<td></td>
<td></td>
<td>110.82</td>
<td>8</td>
<td>.0001</td>
<td>.49</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>110.82</td>
<td>8</td>
<td>.0001</td>
<td>.49</td>
</tr>
<tr>
<td>Syntax</td>
<td>1</td>
<td>.80</td>
<td>.64</td>
<td>113.18</td>
<td>10</td>
<td>.0001</td>
<td>.49</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>2</td>
<td>.17</td>
<td>.03</td>
<td>3.36</td>
<td>4</td>
<td>NS</td>
<td>.01</td>
</tr>
<tr>
<td>Cloze</td>
<td></td>
<td></td>
<td></td>
<td>113.18</td>
<td>10</td>
<td>.0001</td>
<td>.49</td>
</tr>
<tr>
<td>Coherent</td>
<td></td>
<td></td>
<td></td>
<td>3.36</td>
<td>4</td>
<td>NS</td>
<td>.01</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td>113.18</td>
<td>10</td>
<td>.0001</td>
<td>.50</td>
</tr>
</tbody>
</table>
statistics did not change markedly in the second and third analyses. In other words, the addition of coherent and random task performance did not result in stronger between-sets correlations nor in the extraction of a significant amount of variance by another canonical root. Practically speaking, further explanatory power was not gained by the inclusion of random and coherent task performance in the canonical correlation model. On the other hand, their presence in the model did not result in a loss of information. Inspection of the magnitude of the structure coefficients in Table 4.7 confirms that while syntax, vocabulary, and cloze tests were the most prominent variables associated with the canonical variate of the left set, moderately-high, positive loadings were also found for coherent and random task scores.

These results imply that the most important variables defining the canonical variate of the left set were cloze, vocabulary, and syntax in that order. The canonical variate of the right set was defined in terms of both recall and recognition. A substantial correlation between the pair of canonical variates, as well as considerable overlap of information between sets, suggest that skill in reading comprehension is highly predictable from indices of second-language cloze test performance, proficiency in vocabulary, and knowledge of syntax. As might be expected, word identification skills contributed little to the explanation of variance in the reading comprehension factor.

Follow-up Procedures to the Canonical Correlations

The advantage of employing canonical variate analysis was that, via this technique, the nature and strength of the relationship between proficiency in a foreign language and reading comprehension in that language
TABLE 4.7

CANONICAL WEIGHTS AND STRUCTURE COEFFICIENTS OF LEFT AND RIGHT SET VARIABLES ASSOCIATED WITH THE SIGNIFICANT PAIR OF CANONICAL VARIATES (ROOTS) IN EACH ANALYSIS

<table>
<thead>
<tr>
<th>CANONICAL ANALYSIS</th>
<th>Original Variables</th>
<th>Standardized Canonical Weights</th>
<th>Structure Coefficients</th>
<th>Original Variables</th>
<th>Standardized Canonical Weights</th>
<th>Structure Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Syntax</td>
<td>.270</td>
<td>.875</td>
<td>Recall</td>
<td>.754</td>
<td>.957</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>.312</td>
<td>.905</td>
<td>Recognition</td>
<td>.353</td>
<td>.788</td>
</tr>
<tr>
<td></td>
<td>Cloze</td>
<td>.514</td>
<td>.939</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Syntax</td>
<td>.190</td>
<td>.869</td>
<td>Recall</td>
<td>.767</td>
<td>.961</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>.301</td>
<td>.898</td>
<td>Recognition</td>
<td>.337</td>
<td>.779</td>
</tr>
<tr>
<td></td>
<td>Cloze</td>
<td>.500</td>
<td>.933</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coherent</td>
<td>.147</td>
<td>.664</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Syntax</td>
<td>.194</td>
<td>.867</td>
<td>Recall</td>
<td>.755</td>
<td>.958</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>.305</td>
<td>.896</td>
<td>Recognition</td>
<td>.351</td>
<td>.787</td>
</tr>
<tr>
<td></td>
<td>Close</td>
<td>.488</td>
<td>.930</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Coherent</td>
<td>.100</td>
<td>.661</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Random</td>
<td>.090</td>
<td>.430</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
could be viewed from a multivariate perspective. One disadvantage, how­ever, was the inability to exercise control over extraneous variables such as university or class. Although theoretically possible, the in­clusion of those variables in the canonical correlation would have served only to complicate the interpretation of the canonical variates and the assessment of the relationships. Since, however, sample effects were allowed to operate freely in the canonical, the extent to which they may have distorted the findings of the correlational analyses was unclear. Therefore, multiple regression analyses were performed to determine whether, beyond sample effects, the variables used as predictors in the canonical analyses still contributed significantly to the explanation of variance in the dependent measures.

The teacher-class variable, described in the first section of this chapter, was chosen to represent sample effects. This variable incor­porated more levels and, therefore, implied greater variability than the university variable. Given that teacher-class was a categorical variable consisting of eight levels, the dummy codes assigned to designate each subject's group membership resulted in the creation of seven vectors. The original predictor variables (syntax, vocabulary, cloze, random and coherent tasks), along with the transformed teacher-class variable were regressed on measures of recall and recognition in separate analyses. A total of ten stepwise multiple regressions, five for each criterion measure, were performed using the subprogram BMDP2R in Biomedical Compu­ter Programs (Dixon, 1977). The procedures for each regression were the same. The seven vectors associated with the dummy variable were entered into the model before any of the other predictor variables. Then, in
turn, each of the five remaining variables was entered into the regression equation at the eighth step in separate analyses. In this way, the proportion of explained variance attributable to the constituents of teacher-class could be aggregated and their effects isolated in order to examine the increase in explanatory power provided by each additional variable. The results presented in Table 4.8 help to clarify these procedures.

Summarized in this table are the multiple correlations and the squared multiple correlations for each variable. The multiple correlation between recall and teacher-class (considered as an aggregate) was .476. Squaring that coefficient, approximately 23% of the variance in recall was explained by the teacher-class variable. Multiple R's and $R^2$'s associated with the other predictor variables were those obtained as each measure was entered at the eighth step. In addition, for each of these variables, the increment in the squared multiple correlation from the seventh to the eighth step is also presented. The results show that the greatest increase in $R^2$ obtained when the cloze test was entered at the eighth step, followed by the increment due to vocabulary in the same position, then syntax, and so forth.

Having tabulated these findings, the next objective was to test for the significance of the increments in $R^2$ for each respective variable. This was accomplished by calculating $F$-ratios according to the formula suggested by Kerlinger (1973, pp. 70-72). The results of the $F$ tests showed that the value of the multiple $R^2$ obtained at the seventh step increased significantly with the addition of each of the other variables in the next step. Statistical significance, however, does not always
### TABLE 4.8

**SUMMARY OF TESTS OF SIGNIFICANCE FOR INCREMENTS IN SQUARED MULTIPLE CORRELATIONS BETWEEN THE PREDICTOR VARIABLES AND RECALL AFTER PARTIALLING OUT TEACHER-CLASS EFFECTS**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Multiple R</th>
<th>Multiple R Squared</th>
<th>Increment in $R^2$</th>
<th>F (df=1,99)</th>
<th>p &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher-Class (Aggregate)</td>
<td>.476</td>
<td>.226</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloze</td>
<td>.746</td>
<td>.556</td>
<td>.330</td>
<td>73.4763</td>
<td>.0001</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.708</td>
<td>.501</td>
<td>.275</td>
<td>54.4384</td>
<td>.0001</td>
</tr>
<tr>
<td>Syntax</td>
<td>.683</td>
<td>.467</td>
<td>.241</td>
<td>44.6540</td>
<td>.0001</td>
</tr>
<tr>
<td>Coherent</td>
<td>.582</td>
<td>.339</td>
<td>.113</td>
<td>16.8796</td>
<td>.001</td>
</tr>
<tr>
<td>Random</td>
<td>.503</td>
<td>.253</td>
<td>.026</td>
<td>3.4699</td>
<td>.05</td>
</tr>
</tbody>
</table>
imply substantive importance. While cloze, vocabulary, syntax, and coherent task explained important increments of variance in recall, the increment associated with the random task was trivial. Nevertheless, the F tests confirmed that after partialling out sample effects, a significant amount of explanatory power was gained each time a measure from the battery of language tests was entered into the model.

Similar results were found for the analyses conducted on recognition. These are summarized in Table 4.9. Approximately 17% of the variance in recognition was explained through the aggregate. Once again, F tests revealed that considerable explanatory power would have been lost without the contributions of vocabulary, cloze, and syntax, in that order. Although the increments in $R^2$ associated with the coherent and random tasks were statistically significant, neither were substantial.

The results of the multiple regression analyses, designed to control for sample effects, were reassuring. The increase in the squared multiple correlations reported for the measures used as predictors in the canonical (cloze, vocabulary, and syntax, for the most part) indicated that all three contributed substantially to the explanation of variance in both recall and recognition measures after class effects had been accounted for. Given the prominence of these three variables throughout all of the analyses, it is doubtful that the effects of class would have suppressed their importance in the canonical. This assumption is supported by the fact that the random and coherent tasks did not alter the results or interpretation of the canonical analyses, and statistically speaking, their presence in the regression model added significantly to the $R^2$ obtained for the aggregate teacher-class. Had class been included...
<table>
<thead>
<tr>
<th>Variable</th>
<th>Multiple R</th>
<th>Multiple R Squared</th>
<th>Increment in $R^2$</th>
<th>$F$ (df=1,99)</th>
<th>$p &lt;$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher-Class (Aggregate)</td>
<td>.407</td>
<td>.165</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>.666</td>
<td>.443</td>
<td>.278</td>
<td>49.4203</td>
<td>.0001</td>
</tr>
<tr>
<td>Cloze</td>
<td>.655</td>
<td>.429</td>
<td>.263</td>
<td>45.6523</td>
<td>.0001</td>
</tr>
<tr>
<td>Syntax</td>
<td>.619</td>
<td>.384</td>
<td>.218</td>
<td>35.0232</td>
<td>.0001</td>
</tr>
<tr>
<td>Coherent</td>
<td>.469</td>
<td>.220</td>
<td>.055</td>
<td>6.9571</td>
<td>.01</td>
</tr>
<tr>
<td>Random</td>
<td>.460</td>
<td>.212</td>
<td>.047</td>
<td>5.8546</td>
<td>.05</td>
</tr>
</tbody>
</table>
in the canonical, one might speculate that the content of the first pair of canonical variates would not have changed markedly, but the second canonical root may have been significant, in which case, the predictor set variate would, perhaps, have been defined by the constituents of teacher-class.

Based on the results of the multiple regression analyses, the interpretations of the canonical analyses maintained their relevance. When students of a second language are faced with text written in a foreign language, almost half of the variance in reading comprehension can be explained by the individual's ability to infer meaning as well as his/her proficiency in the vocabulary and syntax of the second language.

Summary

This chapter reported the results of the analyses conducted in this study. These included 1) the analysis of a word-boundary task experiment designed to investigate the effects of specific variables on performance, and to determine the extent to which performance on that instrument reflects context utilization, and 2) a description of the relationships found between a set of language proficiency measures and a set of reading comprehension measures. Discussions followed the presentation of the findings. A summary of the procedures and results of this study are presented in Chapter V along with implications for future research.
CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

Overview

The assumptions underlying this research were that prior knowledge is the basis for all comprehension, and that readers use prior knowledge strategically to construct meaning for text by integrating their knowledge of language with their knowledge of the world in general. Knowledge of a language constitutes only a part of the background knowledge an individual can bring to bear on a particular reading. The less a reader understands about the language in which a text is written, however, the more manipulation of all available information sources is required of that reader to comprehend successfully. Against this background, the present study sought to examine the relationship between measures of reading comprehension of texts written in Italian and a set of second-language proficiency measures including knowledge of Italian vocabulary and syntax, the ability to infer meaning in a cloze passage, and measures of context utilization as defined by performance on a word-boundary task. The inclusion of context utilization within the set of proficiency factors, however, was contingent upon the conclusions drawn from another line of inquiry that investigated the methodology and construct validity of the instrument employed to measure context utilization. In the following sections, the procedures of the study are summarized, the findings are discussed, and recommendations for further research are offered.
Summary of Procedures

All instruments employed in this study were either developed or adapted by the researcher to measure the constructs of interest. Pilot tests were conducted to assess instrument reliabilities and to determine the appropriateness of the materials and procedures of the study. Reading comprehension was defined by measures of recall and recognition of two passages written in the target language. Proficiency in Italian was assessed by means of vocabulary and syntax tests, and performance on a cloze task. In addition, two measures were obtained from a word-boundary task that reflected performance on coherent and random versions of a passage. Performance on the word-boundary task was also a function of the presence or absence of contextual clueing and sentential boundary markers, as well as the presentation sequence of the random and coherent tasks. The materials prepared for the word-boundary task, therefore provided for eight distinct performance conditions. One form of the word-boundary task was placed in a test booklet along with the vocabulary test and the tests accompanying one of the reading passages. The remaining tests of reading comprehension, Italian syntax, and cloze task comprised another test booklet. Neither the order of instruments within each test booklet, nor the presentation sequence of the booklets was varied; thus, all copies of each test booklet were the same with the exception of the WBT form which varied along experimental lines. Testing procedures were identical for participants in the study.

All subjects in the study consented to participate willingly. They were 114 second-language students enrolled in intermediate courses of Italian at three state universities. Data were collected on two days
during hours regularly scheduled for class sessions. Test booklets containing the different WBT forms were distributed evenly to groups, and distributed randomly to subjects within each group. Each subject was present at both testing sessions to complete data on all instruments.

The data were analyzed via computer facilities. Preliminary analyses included reliability reassessments of the instruments, transformation of recall and recognition raw scores to standardized Z scores, a descriptive work-up of sample characteristics, and multivariate and univariate analyses of variance among university and class samples. Differences among universities and classes were found on each dependent measure; therefore, these variables were articulated in subsequent analyses.

Two analyses of variance were performed on the data from the WBT experiment utilizing a repeated-measures design. Levels of the university variable were controlled in one of the analyses. The results of both ANOVAs were essentially the same. Significant effects were followed-up using Scheffé and Tukey post-hoc procedures. The methodological artifacts associated with the WBT were discussed in light of the findings from the experiment, and those outcomes also served as a basis for evaluating the construct validity of the WBT as a measure of context utilization.

An examination of the relationship between reading comprehension and second-language proficiency was conducted using multivariate correlational procedures. Initially, data were grouped into two sets. The "dependent" set included measures of recall and recognition; the "independent" or predictor set was comprised of vocabulary, syntax, and cloze task measures. A canonical variate analysis was performed to determine the
nature and strength of the relationship between the two factors or composites. The canonical correlation reported from this analysis remained relatively unchanged by the inclusion of random and coherent WBT performances within the independent variable set. Multiple regression analyses were performed on recall and recognition separately, to test for increments in $R^2$ due to each predictor variable after partialling out extraneous teacher-class effects.

Summary of Findings

Review of Experimental Outcomes

Statistical control was exercised on four factors incorporated in the design of the word-boundary task. These factors were contextual clueing, boundary markers, coherent vs. random task performance, and the presentation sequence of the two tasks. The results of the WBT experiment revealed that performance on coherent and random tasks was affected by the presentation sequence of the two tasks. A significant interaction was also found among three of the variables, independent of task sequence. In addition, a strong main effect for the task factor emerged. The interpretation of the task main effect was postponed to follow up the interaction effects; in both cases, however, post-hoc inspections confirmed the persistence of the task factor. The most reliable finding from these analyses was that context, present in the coherent passage, facilitated word identification. This overall finding was compatible with reports of previous WBT research, yet, other outcomes of the analyses performed in this study did not support former claims of the facilitative effects of sentential boundary markers on WBT performance (Klein, Klein, and
Hildum, 1974) and the absence of effects for a contextual framework statement (Schwartz, 1977).

Klein et al. (1974) found that markers placed at the end of sentences in coherent passages facilitated WBT performance. A similar effect was not observed in this study. Although performance was somewhat improved by the presence of boundary markers in the coherent task, no particular influence of that variable was identified. One explanation for this discrepancy is the difference in average sentence length between the two studies. In their experiments, Klein et al. found significant effects for sentence markers in shorter sentences ($\bar{x} = 6$ words) that were not found in longer sentences ($\bar{x} = 19$ words). The mean sentence length of the coherent passage used in this study was approximately 9 words. It is possible that increased sentence length, combined with the other variables operating in this experiment, but not in the Klein et al. study, accounted for the differential effects.

Similarly, no main effect was found in this study for contextual clueing, and the context clue by boundary markers by task interaction also revealed no strong effects due to that variable in particular. Verbal contextual clueing did play a prominent role in interaction with the university variable, however, when the university factor was included in the analysis of the WBT data. More specifically, contextual clueing enhanced the overall WBT performance of subjects in one university sample. Because the mean scores for this sample on the majority of language proficiency and reading comprehension measures were significantly lower than the other groups' means, contextual clueing was interpreted as having a greater facilitative effect on the lower proficiency group. This
explanation is based upon an interactive-compensatory model of reading, and is supported by first-language reports that have shown less fluent readers to be more reliant upon context than fluent readers (see Stanovich, 1980).

The conditional facilitative effect for verbal contextual clueing in this study, however, was inconsistent with Schwartz's (197?) findings of no differences in performance attributable to a framework statement. Schwartz explained, however, that effects may have emerged, had the framework statement been furnished for a longer time segment prior to WBT performance. His suggestions were observed in this study, and contributed perhaps, in part, to the effects found with contextual clueing. Another plausible reason for the difference between Schwartz's results and those reported by this study might have been the difficulty of the passages. Schwartz argued that the context of the passage used in his first experiment was perhaps detected easily, without the aid of a framework statement. The absence of a main effect for context clueing in this study suggests a similar conclusion. The interaction of context clue with the university factor, however, implied that for some individuals, perhaps those less proficient in the language, a contextual clue facilitated the construction or identification of meaning. This point buttresses the interpretation offered for the context clue by university interaction that was found in this analysis of WBT performance.

The result that differed most markedly from previous reports of WBT experiments was the significant task by task sequence interaction. This difference among findings was also the most easily explained. Only one previous study (Klein, Klein, and Bertino, 1974) incorporated strictly
random and coherent tasks, and those tasks were not presented back to back as was done in this study. Instead, four unrelated passages were prepared in a WBT format with both a coherent and a random version. Subjects received two coherent and two random tasks, each from a different prose selection. No order effects were found, but no subject received both coherent and random versions of the same passage, as was the procedure for all subjects in this study. In a sense, the design of this study provided a more stringent and stronger test of the effects of context on word identification than did previous studies, since subjects were presented with versions of the same passage differing only with respect to the order imposed by contextual constraints. Taking this into consideration, the most surprising finding in this study was, perhaps, the absence of a main effect for task sequence. Indeed, that effect did not even approach significance in either WBT analysis (including university: $F(1,90) = .30; p < .59$ / excluding university: $F(1,106) = .33; p < .59$). Moreover, the task by task sequence interaction only hinted at a practice effect that was not substantiated by follow-up procedures.

These findings provide strong evidence for the effects of context on WBT performance.

It is possible that the discrepancies among findings presented above were due, in part, to the fact that this study examined WBT performance in a second language, whereas all previous studies were conducted in English with native speakers. This possibility remains open to question. It appears more likely, however, that comparable effects were not found, because of the different research methodologies employed. The dissimilarities among previous WBT findings and those reported here were not
dramatic; moreover, strong effects for context were manifested throughout. But, across studies, the instability of the effects of other factors examined, as well as their differential effects upon random and coherent tasks, jeopardizes somewhat the stability or validity of measures of context utilization. More specifically, if WBT performance is affected to a great extent by changes in methodology, then it appears that the factor structure of the measures obtained and any measure derived from the originals will be altered accordingly. The interactions that were demonstrated in this study, for example, did not dispel the salient effect of context on word identification; however, they did affect random and coherent task performance differently and therefore obscured the meaning of context utilization measures based upon ratios or differences between the scores on the two tasks.

The validity of the WBT as a measure of context utilization appears, at least partially, to be a function of the methodology employed, and, at present, many of the methodological questions associated with the instrument remain unresolved.

Review of Correlational Findings

Three canonical variate analyses were performed to examine the relationship between measures of reading comprehension and second-language proficiency. In each analysis reading comprehension was defined by measures of recall and recognition. The set of language proficiency measures included vocabulary, syntax, the cloze task, and random and coherent WBT measures. (Measures of context utilization were not obtained because of issues discussed in the preceding section). When this complete set of language proficiency factors was analyzed for its
relationship to the set of reading comprehension measures, one highly significant canonical root emerged; the other root possible, however, was non-significant. The maximum correlation between the two variable sets was .80 and the proportion of variance explained in the criterion set composed of reading comprehension measures, given the other variable set, was 50%. Although the canonical variate for reading comprehension reflected high loadings from both recall and recognition measures, the most prominent variables associated with the variate for language proficiency were the cloze task, vocabulary, and syntax measures. Neither coherent nor random WBT scores contributed significantly to the explanation of variance. When only cloze, vocabulary, and syntax measures were included in the predictor set, the canonical correlation was .79 and the percentage of variance shared by the pair of variates was still 48%.

To determine the extent to which extraneous sample effects may have been operating in the canonical analyses, multiple regression follow-up procedures were conducted on recall and recognition measures separately, partialling out class effects. These analyses indicated that, beyond sample effects, a significant amount of variance in both dependent measures was explained by each of the language proficiency factors. The most substantial increments, however, were attributed to the cloze task, vocabulary, and syntax measures. These findings were in accordance with those reported for the canonical analyses. It was therefore concluded that the variance due to class effects would not have distorted the results of the canonical significantly. Thus, approximately half of the variance in reading comprehension was explained through the joint contributions of measures of cloze task performance and vocabulary and syntactic knowledge.
On the one hand, it is evident from these findings that proficiency in the second language accounted for a substantial amount of variance in reading comprehension of passages written in Italian. On the other hand, a proportionate amount of unexplained variance still remained. Knowledge of a language and the ability to use that knowledge are essential factors in reading comprehension, however, other factors, cognitive as well as affective perhaps, are also involved in the ability to comprehend second-language materials.

**Recommendations for Further Research**

The generalizability of all research findings is dependent upon the sample drawn and the reliability of the instruments employed. Questions addressed in this study should therefore be asked of data obtained from students of other second languages, utilizing if possible, standardized measures of reading comprehension and second-language proficiency.

The large proportion of unexplained variance in reading comprehension of a second language reported by this study, allows for further examination of factors associated with that construct. The relationship between reading comprehension and second-language proficiency factors might be pursued by including affective measures such as motivational or personality factors within the predictor variable set. It might also prove interesting to incorporate measures of first-language proficiency and reading comprehension within the battery of predictive measures, as well as measures of verbal and nonverbal problem-solving skills.

A review of the literature in reading comprehension and strong correlations between that construct and cloze task performance suggest that reading comprehension entails skill in integrating various sources of
information. Perhaps training of second-language students in strategic processing skills would enhance the comprehension abilities of those students. The cloze task as well as the word-boundary task might be employed effectively for that purpose (see Bialystok, 1979 and Schwartz, 1980). An investigation of the effects of such training would provide important information for educators and researchers alike. The cloze task and the word-boundary task must be tapping some of the same strategies, but different ones as well; they might therefore be compared for their utility as strategic training devices with second-language students of various proficiency levels.

To attain a better understanding of the strategies employed during word-boundary task performance, subjects should be tested individually, rather than in groups, on random and/or coherent materials. Subjects might be allowed as much time as desired to complete the tasks, and the times recorded for each individual might be correlated with performance scores. Both the number of correctly identified words, as well as the number and types of errors should afford valuable information concerning comprehension and language problems of second-language students. Observing WBT performance on an individual basis might also provide insights into factors involved in word-boundary task performance.

One factor that should be given attention, if questions concerning the validity of WBT measures are to be pursued, is the time variable. Controlling for WBT performance time might be a way of determining whether, for second-language subjects, the WBT is a "pure" speed test (as it probably is for most native-language subjects) or a "power" test. If, for example, variance in the number of errors or correctly identified
words is still observable from data of second-language subjects given no time constraints, then the WBT would appear to be a power test for some or all of those subjects. In that case, more valid WBT responses might be obtained by setting realistic time limits. Appropriate time limits for passages of varying lengths could be established systematically, by controlling time and length and examining performance under the varied conditions.

The word-boundary task has demonstrated its value as a research tool for investigating the effects of specific factors on word identification, and for examining their effects in light of theories and models of reading. New knowledge about the reading comprehension process might be gained by continued investigation of word-boundary task performance, under varied conditions, and with students and native speakers of different languages and abilities.
BIBLIOGRAPHY


Anderson, R.C., & Freebody, P. Vocabulary knowledge. (Technical Report No. 136.) University of Illinois at Urbana-Champaign, 1979. (ED 177 480)


Botzum, W.A. A factorial study of reasoning and closure factors. Psychometrika, 1951, 16, 361-386.


Davis, F.B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 2, 185-197.


Labarca, A. The effects of time of presentation and type of diagrammatic organizer on recall measures of reading comprehension in beginning college Spanish. (Doctoral Dissertation, The Ohio State University, Columbus, Ohio, 1979).


Mueller, G.A. The effects of a contextual visual on recall measures of listening comprehension in beginning German. (Doctoral Dissertation, The Ohio State University, Columbus, Ohio, 1979).


Neufeld, J.K. The effects of new word density on measures of reading comprehension in university English as a second language students. (Doctoral Dissertation, The Ohio State University, Columbus, Ohio, 1980).


Omaggio, A.C. The effects of selected pictorial contexts on measures of reading comprehension in beginning college French. (Doctoral Dissertation, The Ohio State University, Columbus, Ohio, 1977).


Phillips, J.K. A study of the applicability of task analysis methodology and learning hierarchies to second language reading. (Doctoral Dissertation, The Ohio State University, Columbus, Ohio, 1974).


Stotsky, S.L. Vocabulary: A major factor in reading comprehension. (Study prepared at Harvard University, 1976). (ED 142 946)


Thurstone, L.L. Note on reanalysis of Davis' reading tests. *Psychometrika*, 1946, 11, 185-188.


APPENDIX A

MATERIALS FOR MEASURING
READING COMPREHENSION
Please read the following passage about marriage in Italy

Anche se il matrimonio è sempre stato soggetto a critiche, questa visione pessimistica della vita coniugale non ha mai avuto troppe conseguenze pratiche. In passato il numero dei matrimoni è sempre aumentato con il crescere della popolazione. Dagli anni sessanta però l'istituzione del matrimonio si è trovata in crisi insieme a tanti altri valori. In Italia ci si sposa sempre meno. Questo non significa che non ci s'innamora più. Anzi, anche dall'introduzione del divorzio le statistiche mostrano che sempre più giovani scelgono di vivere come marito e moglie senza sposarsi.

Molti sociologi e psicologi spiegano che il numero di matrimoni è ridotto a causa di un certo cambiamento di mentalità. Una volta il matrimonio era il "coronamento dell'amore" per i romantici, un sacramento per i religiosi e per quasi tutti un'alleanza eterna dedicata alla procreazione dei figli. Il matrimonio fissava i ruoli della donna e dell'uomo. La moglie allevava i bambini e lavava i piatti in casa mentre il marito lavorava. In passato le cose funzionavano perché i ruoli erano ben definiti: l'uomo comandava e la donna ubbidiva. Oggi però il concetto del matrimonio è cambiato. La donna e l'uomo cercano una certa parità nei rapporti. Non si vuole perdere la propria identità conformandosi al comportamento tradizionale che il matrimonio porta con sé.

Ci sono probabilmente molte altre cause per il declino dei matrimoni come, per esempio, i problemi economici. Ma una cosa è certa. Il matrimonio del futuro non sarà più il matrimonio classico di una volta.
Please write a summary in English of the passage you have just read. In complete sentences provide as much information as you can remember from the reading.
Circle the letter of the response that makes each statement true, according to the passage you have just read.

1. The decline in the number of marriages is primarily due to
   a) economic problems
   b) a breakdown in moral values
   c) the introduction of divorce
   d) a change in the way of thinking

2. The marriage crisis in Italy began
   a) when divorce was legalized
   b) with the decline in Church attendance
   c) during the 1960's
   d) with the advent of the women's liberation movement

3. At one time marriages in Italy worked because
   a) the roles of men and women were clearly defined
   b) couples couldn't live together legally outside of marriage
   c) the financial situation was such that wives didn't have to work outside the home
   d) people had a deeper respect for the sacrament of marriage

4. Today in Italy many young people consider marriage
   a) a threat to their individuality
   b) a romantic illusion
   c) only when and if they intend to have children
   d) only after first living together outside of marriage

5. In recent years the institution of marriage has been criticized because
   a) for the most part, the disadvantages of married life tend to outweigh the benefits
   b) it forces people to conform to a traditional mode of behaviour
   c) couples can hardly afford to have children any longer
   d) the classic ideal of an eternal bond to one person is no longer realistic

PLEASE DO NOT TURN THE PAGE UNTIL A SIGNAL IS GIVEN
Please read the following passage about snow skiing.

In estate la maggior parte della gente va al mare. C'è però un piccolo gruppo di persone che ama lo sci, e per queste, fortunatamente, esistono ora in Italia numerose scuole di sci estivo. Lo sci estivo non è molto diverso da quello invernale. Certo, durante i mesi caldi la qualità della neve non è della migliore. È possibile sciare da maggio a settembre ma non c'è quella bella neve d'inverno. È vero che sciare con la neve pesante è più difficile proprio perché l'esecuzione di ogni movimento deve essere perfetta per riuscire bene, ma durante i mesi estivi il progresso nella tecnica dello sci è più rapido.

Lo sci estivo è più organizzato di quello invernale. Chi lo pratica deve farlo in una scuola sotto la guida di maestri capaci, lontano dalle distrazioni caratteristiche dell'inverno. Il risultato è che si impara meglio e più rapidamente.

La vita in una scuola di sci estivo dura una settimana ed è molto disciplinata. In genere, gli studenti si svegliano presto la mattina e vanno a letto altrettanto presto la sera. Dopo un'abbondante prima colazione e qualche esercizio preparatorio seguono tre o quattro ore di lezioni in piccole classi sotto la guida di esperti maestri. Il pomeriggio quando la neve è troppo pesante per sciare, si riposano e prendono il sole. Il sole di montagna brucia e abbronzà di più di quello del mare ed è necessario proteggere la pelle. Quindi, oltre alla normale attrezzatura da sci, ci vuole una buona riserva di creme protettive.

Tutto sommato, i benefici dello sci estivo includono una vita sana e sportiva al sole ed all'aria aperta, via dalle spiagge affollate, dal caldo ossessivo e dall'asfalto delle città.
Please write a summary in English of the passage you have just read. In complete sentences provide as much information as you can remember from the reading.
Circle the letter of the response that makes each statement true, according to the passage you have just read.

1. Only a few people
   a) go mountain skiing during the summer
   b) can withstand the excessive heat of the mountains in summer
   c) have the skills necessary to maneuver themselves on skies in the summer snow
   d) prefer summer skiing to the more difficult winter skiing

2. Those who practice summer skiing
   a) generally prefer it to winter sports
   b) must be very experienced winter skiers
   c) must do so under the direction of capable instructors
   d) must exercise caution against the distractions characteristic of the summer months

3. The schools for summer skiing offer classes
   a) each day for one week
   b) three or four days per week during May or September
   c) one day a week for the period from May to September
   d) each day for three or four weeks

4. Included in the daily routine of the summer sessions at the schools is a
   a) light breakfast
   b) rest period in the afternoon
   c) well organized exercise period from 3-4 in the afternoon
   d) class that deals with health topics (diet, exercise, etc.)

5. Since the quality and texture of the snow differs from season to season, during the summer
   a) it is advisable to ski in small groups
   b) the skiers' every movement must be carefully executed to prevent slipping
   c) one can improve his/her skiing skill more quickly
   d) special ski equipment must be worn

Thank you for your participation in this study
1. Even though marriage has always been a subject of criticism
2. this pessimistic view of married life has never had too many practical consequences
3. In the past, the number of marriages has always increased with an increase in the population
4. Since the '60's, however, the institution of marriage has found itself in crisis
5. along with many other values
6. In Italy people marry less and less
7. This does not mean that they no longer fall in love
8. On the contrary, even since the introduction of divorce, statistics show
9. that more people are choosing to live together as man and wife, outside of marriage
10. Many sociologists and psychologists explain
11. that the number of marriages is reduced
12. due to a change in the way of thinking
13. Once marriage was the "crowning of love" for the romantics
14. a sacrament for the religious
15. and for almost everyone, an eternal bond
16. intended for procreation
17. Marriage set the roles for men and women
18. The wife raised the children
19. and washed the dishes
20. while the husband worked
21. In the past things worked out
22. because the roles were well defined
23. The man ruled
24. and the woman obeyed
25. Today, however, the concept of marriage has changed
26. Men and women look for a certain equality in their relationships
27. One doesn't want to lose his/her identity
28. by conforming to the traditional behaviour
29. that marriage brings with it
30. There are probably many other reasons for the decline in marriage
31. for example the economy
32. But one thing is certain
33. Marriage of the future will no longer be the classic marriage of one time
1. In summer the majority of people go to the sea
2. There is, however, a small group of people who love skiing
3. and for those, fortunately, there exist now in Italy numerous
   schools for summer skiing
4. Summer skiing is not very different from winter skiing
5. Certainly, during the hot months the quality of the snow is not
   the best
6. It is possible to ski from May to September
7. but without the beautiful winter snow
8. It is true that skiing in heavy snow is more difficult
9. precisely because every movement must be perfectly executed
10. in order to succeed well
11. but during the summer months, progress in skiing skill is more rapid
12. Summer skiing is more organized than winter skiing
13. Those who practice it must do so in a school
14. under the supervision of capable teachers
15. far from the distractions characteristic of winter
16. The result is that one learns better and
17. more rapidly
18. Life in a summer ski school lasts one week
19. and is very disciplined
20. In general, the students get up early in the morning
21. and go to bed just as early in the evening
22. After a filling breakfast
23. and some warm-up exercises
24. follow three or four hours of lessons in small classes
25. taught by expert teachers
26. In the afternoon, when the snow is too heavy for skiing
27. they rest
28. and take in the sunshine
29. The mountain sun burns
30. and tans
31. more than the sun near the sea
32. and it is necessary to protect the skin
33. Therefore, besides the normal ski equipment
34. one should have a good supply of sun-blocking cream
35. To summarize, the benefits of summer skiing include
36. a healthy life
37. in the fresh air
38. far from crowded beaches,
39. excessive heat,
40. and the city streets
PASSAGE DIFFICULTY RATING SCALE

On the following five-point scale please rate the reading passage in terms of its comprehension difficulty.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>very easy</td>
<td>fairly easy</td>
<td>challenging but I understood</td>
<td>rather difficult</td>
<td>very difficult</td>
</tr>
</tbody>
</table>
Fill in each blank with the appropriate English word that best completes the meaning of the following paragraph. In addition, if you know the Italian equivalent, write it above the English word. Do not choose an English word simply because you know the Italian version. It is important that the paragraph make sense.

Franco frequenta il secondo anno di Legge o almeno così dice il suo libretto universitario. In realtà, lui passa quasi l'intera ___________ al caffè. Arriva al caffè ogni ___________ alle otto in punto per ___________ un cappuccino e una pasta. Due ___________ dopo torna per bere un caffè ___________ e poi a mezzogiorno per un ___________. Verso le due lui ritorna per un ___________ caffè ristretto tanto per non dormire ___________ pomeriggio. Alle nove di sera si trova ___________ al caffè per un liquorino digestivo e ___________ una bella discussione politica o una ___________ a carte con gli amici. Al ___________ tutti lo conoscono. La gente pensa che ___________ è un tipo simpatico e anche ___________ allegro.

THANK YOU VERY MUCH FOR YOUR PARTICIPATION TODAY
Circle the letter of the item that best completes the meaning of each sentence.

* 1 zingaro è 2 vagabondo.
   a) gli  a) un
   b) li'  b) un'
   c) il   c) una
   d) lo   d) uno

* Le ragazze sono 3.
   a) belli
   b) male
   c) presenti
   d) intelligente

* Giorgio dà i libri a Paola e lei 4 accetta.
   a) la
   b) li
   c) lo
   d) le

* Ecco 5 frutta e 6 biscotto.
   a) di     a) qualche
   b) molta  b) alcune
   c) poco   c) dello
   d) un po' d) un po'

* 7 alzerai presto domani.
   a) si
   b) tu
   c) ti
   d) ci

* Il padre 8 donna è morto.
   a) di
   b) da
   c) della
   d) dalla

* Gianni cerca 9 10 giacca.
   a) il     a) sua
   b) la     b) sue
   c) le     c) suo
   d) no article d) suoi

YOU MAY TURN THE PAGE AND CONTINUE
* l'___11___ sono ___12___ simpatici.
  a) animali  a) molti
  b) signori  b) mai
  c) studenti  c) non
  d) zii     d) molto

* La ragazza ti ___13___.
  a) piaci
  b) piaciuta
  c) piace
  d) piacciono

* Teresa ed io ___14___ insieme.
  a) studierò
  b) studierete
  c) studieranno
  d) studieremo

* ___15___ ___16___ ho mai vedute.
  a) le    a) le
  b) lo    b) lo
  c) non   c) non
  d) gli   d) gli

* ___17___ sono già partiti.
  a) io
  b) Maria e Lidia
  c) Tu e Giuseppe
  d) loro

* Mario, ___18___ la lezione ora!
  a) finisce
  b) finisci
  c) finisca
  d) finire

* Luigi è un ___19___ ragazzo ___20___.
  a) bello      a) bene
  b) italiano   b) sincere
  c) bravo      c) meglio
  d) intelligente d) intelligente

* Si ___21___ divertite.
  a) sono
  b) siete
  c) hanno
  d) sei

YOU MAY TURN THE PAGE AND CONTINUE
22 madre è sempre allegra.

a) la          a) tuo
b) le          b) tua
c) l'           c) tue
d) no article   d) tu

24 Loro si   ora.

a) scusano
b) partono
c) escono
d) camminano

25 26 denaro e voi avete 27 28.

a) la          a) loro          a) i          a) vostro
b) le          b) vostre        b) il         b) vostri
b) vostre

29 non 30.

a) mi          a) sè
b) io          b) tu
b) tu

31 sono piaciuti.

a) le
b) loro
b) loro

c) li

d) lo

32 ieri.

a) arrivato
b) arrivati
b) arrivati

c) studiato
c) studiato
d) studiati
d) studiati

33 studiare.

a) di
b) da
b) da
c) a
c) a

d) no preposition
d) no preposition

34 vostri cugini

a) a
b) ai
b) ai
c) agli
c) agli
d) i
d) i
Il vecchio _______ continuamente del tempo passato.
   a) parlò
   b) ha parlato
   c) parleranno
   d) parlava

Quando li vedo _______ parlano sempre del lavoro.
   a) mi
   b) me
   c) lo
   d) li

Devo _______ partire stasera.
   a) di
   b) da
   c) a
   d) no preposition

Marcella parla molto _______.
   a) chiara
   b) chiare
   c) chiaramente
   d) chiarissima

Quando Anna vede Enzo _______ dà la mano.
   a) le
   b) gliene
   c) lo
   d) gli

A _______ parlate?
   a) che
   b) cui
   c) chi
   d) ciò
Circle the letter of the item that is closest in meaning to the Italian word.

1. fra a) through  b) for  c) between  d) beside
2. albero a) hotel  b) suit  c) corner  d) tree
3. perdere a) to clean  b) to take  c) to rain  d) to lose
4. quasi a) almost  b) some  c) any  d) at least
5. prezzo a) piece  b) price  c) quickly  d) dinner
6. troppo a) more  b) very  c) too much  d) everything
7. marito a) husband  b) sweater  c) market  d) pencil
8. aiutare a) to hear  b) to be bored  c) to help  d) to exhaust
9. poi a) little  b) however  c) then  d) more
10. uscire a) to leave  b) to go out  c) to dry  d) to use
11. stagione a) season  b) station  c) room  d) reason
12. quanto a) that  b) which  c) when  d) how much
13. venire a) to live  b) to go out  c) to see  d) to come
14. paura a) fear  b) power  c) rain  d) tear
15. gli a) the  b) some  c) to  d) they
16. stanco a) lonely  b) tired  c) slow  d) sad
17. impiegato a) winter  b) raincoat  c) guest  d) clerk
18. ascoltare a) to raise  b) to wait for  c) to hear  d) to listen to
19. uomo a) egg  b) official  c) man  d) work
20. facile a) cold  b) easy  c) fresh  d) useful

YOU MAY TURN THE PAGE AND CONTINUE
21. abbastanza a) without b) enough c) across d) only
22. notizia a) news b) great aunt c) store d) niece
23. orario a) clock b) timetable c) crowd d) arrival
24. mese a) month b) meal c) way d) half
25. scherzare a) to shine b) to dream c) to joke d) to hope
26. ogni a) each b) today c) only d) by now
27. meglio a) less b) mile c) wife d) better
28. biglietto a) ticket b) bottle c) glass d) drink
29. lago a) place b) lake c) wide d) slow
30. appena a) ahead b) rather c) even d) just
31. genitore a) relative b) generation c) parent d) ancestor
32. cambiare a) to carry b) to change c) to wear d) to build
33. secolo a) sequence b) century c) circle d) purpose
34. provare a) to find b) to be able c) to try d) to provide
35. proprio a) proper b) ready c) clean d) really
36. chiedere a) to keep b) to close c) to mean d) to ask
37. fiume a) flower b) river c) outside d) furious
38. alcuno a) every b) none c) some d) at least
39. gita a) glove b) excursion c) finger d) tour
40. chiamare a) to call b) to look for c) to name d) to walk

STOP!

DO NOT TURN THE PAGE UNTIL A SIGNAL IS GIVEN

If you have some blanks remaining in this task, you may turn back to the previous page and complete those items until time is called.
APPENDIX C

WORD-BOUNDARY TASK MATERIALS
This exercise has two parts, but the procedures for each are the same. Your task is to find Italian words within a series of letters that are printed one space apart. When you find a word within the series, indicate its presence by placing a slash at the end of it. There are no extra letters other than those forming words, so a slash indicates the end of one word and the beginning of the next. Since only a brief amount of time will be allotted for the exercise, you are encouraged to mark slashes as quickly as possible, but it is more important that your markings be correct! If you make a mistake, you may cross out the slash and replace it in the appropriate spot. You are to work from left to right, starting at the top and working your way down line by line. No words carry over to a following line, therefore, it will not be necessary to make a slash for the last word on a line. This "free slash" will save you time. Remember to work as quickly and as accurately as possible. Let's try some examples in English and Italian.

unless/c o u n t r y/w a y/p r e s s/d t o/b e s t/o n/t r a v e l
l a/c u c i n a/i t a l i a n a/l e/c o n o s c i u t a/i n/t u t t i/i
p a e s i/d e l/m o n d o

p i z z a/p a r l a/p e r s o n a/l i b e r t à/t e a t r o/m a c c h i n a

p i z z a/p a r l a/p e r s o n a/l i b e r t à/t e a t r o/m a c c h i n a

Are there any questions? The next page you turn to will have a brief message written in English. Please read the message to yourself. It will give you a further instruction about the first task. After reading the message, wait for the signal to begin the first part of the exercise. After you turn the page to the first exercise no questions will be permitted, so if you have further questions regarding the procedure, you may ask them now.

Please turn your booklet to a horizontal position in front of you and wait for the signal to turn the page.
<table>
<thead>
<tr>
<th>FORM</th>
<th>PASSAGE*</th>
<th>INSTRUCTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>GREEN</td>
<td>R</td>
<td>You will be looking for Italian words related to alleviating stress. Two dashes (--) represent a &quot;free slash&quot;. You need not mark a word boundary in these places.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>You will be looking for Italian words that form a paragraph about alleviating stress. Two dashes (--) mark the end of a sentence. You need not put a slash in these places.</td>
</tr>
<tr>
<td>YELLOW</td>
<td>R</td>
<td>You will be looking for Italian words related to alleviating stress.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>You will be looking for words that form a paragraph about alleviating stress.</td>
</tr>
<tr>
<td>BLUE</td>
<td>R</td>
<td>Two dashes (--) represent a &quot;free slash&quot;. You need not mark a word boundary in these places.</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Two dashes (--) mark the end of a sentence. You need not put a slash in these places.</td>
</tr>
<tr>
<td>PINK</td>
<td>R &amp; C</td>
<td>Remember, accuracy is emphasized over speed. If you change your mind about a word, cross out the slash and replace it in the appropriate spot.</td>
</tr>
</tbody>
</table>

* R = random version  
C = coherent version
essere il difendere e superare invece corpo meglio studia esami
rinforzare non anche agli difficoltà consigliabilità per recco
studenti spesso presto verdure e importanti la prima alcuni
soffrono bene questo se sempre ottimi letto essere ed ele per la
periodo infinito tensione medico che devono id dormire i unae
suggerimento molto necessario delle continue e per studiare
molti pause disturbato fine sono omento lo breve vila a tabacco sono
studio essere preparazione andare assolutamente appropriata
tutto è dieta non caffe è l'alcool nutrienti inti i deve per periodo per
sol concentrati frutta pasti per leggeriè

PLEASE DO NOT TURN THE PAGE UNTIL A SIGNAL IS GIVEN
moltisostudentisoffronolatensionedellapreparaziuneagli

esamipersuperaremegliouquestoperiododifficilecco

alcunisuggerimentiidelmedicoprimadituttolostudionon
devesserecontinuoèbeneestudiareperbrevi periodi
concentratiëfarenespessoedellepauseperrinfrescareilcorpo
elamenteuunadietaappropriataèanchenecessariaperchì
studiaipastiëvonoessereeligeralimentenutrientiottimesonole
verduraelafruttainveceilcaffèiltabaccodell'alcoolsono.
sconsigliabiliinfineèmoltoimportanteandarealrettosempre
prestoedormiresoliperononesseredisturbati

PLEASE DO NOT TURN THE PAGE UNTIL A SIGNAL IS GIVEN
molti studenti sono offrono la tensione del loro periodo di preparazione agli esami. Persuadere meglio questo periodo di difficoltà e ecco alcune suggerimenti dal medico - prima di tutto, tutto lo studio deve essere continuo. È bene studiare per brevi periodi e concentrarsi. Fare spesso di pause possono essere utili per rinfrescare il corpo e l'anima. Un'alimentazione appropriata è anche necessaria perché il tuo corpo si prenderà le pause. I pasti devono essere leggeri e nutrienti. Ottimi sono le verdure e la frutta - invece il caffè, l'alcool e il tabacco sono sconsigliabili. Infine è molto importante evitare alcol e tabacco sempre presto ed essere soli per non essere disturbati.

PLEASE DO NOT TURN THE PAGE UNTIL A SIGNAL IS GIVEN
esserei l'diellal alesuperarei inveceecorpaceomegliostudia- -esamirinfrescareanonancheaglidifficilesconsigliabili- -pereccostudiantièsospessoprestoverdureeimportantelaprimaalcunissoffronobenequestosempreottimelettoesseredelaperiaperiodyoinfinetensionemedico- -chedevonoildormireiunessugerimentimoltonecessariadellecontinuo- -epersudiaremoltipausedisturbati- -faresonoemente- -lobrevilaatatabaccosono studioesserepreparazioneandareassolutamenteappropriata tuttoèdietaneoncaffèl'alcoolnutrienti- -ildeveperiodiper soliconcentratiifruttapastiperleggeriè

PLEASE DO NOT TURN THE PAGE UNTIL A SIGNAL IS GIVEN
Please supply the following information by either filling in the blank or checking the appropriate response.

Age: ________
Male ________ Female ________
Undergraduate ________ Graduate ________
Major or minor in Italian: yes ____ no ____
Speak an Italian dialect: yes ____ no ____
Speak or study another language besides English or Italian:
yes ____ no ____
If yes, please specify: __________________

Please do not turn the page until a signal is given.

Today you will be given a series of tasks to complete. Directions will precede each of the tasks and a specific time segment will also be allotted for each. Please start and stop the tasks when the instructions to do so are given.

The first two tasks are based on a short reading passage. After you have read the passage, you will be asked to write an English summary of what you have read. You may not refer to the reading selection when writing your summaries. After you have completed the summary, wait for the signal to complete a short multiple-choice quiz based on the reading selection. Are there any questions?

Wait until you are given instructions to turn the page, then begin reading.
TABLE D.1
DESCRIPTIVE WORK-UP OF PERSONAL DATA FOR TOTAL SAMPLE
AND BY UNIVERSITY

<table>
<thead>
<tr>
<th>Descriptor</th>
<th>Level</th>
<th>University</th>
<th>University</th>
<th>University</th>
<th>Total</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td>Sample</td>
<td>Sample</td>
</tr>
<tr>
<td>Descriptor</td>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>Sex</td>
<td>Males</td>
<td>5</td>
<td>29.4</td>
<td>17</td>
<td>42.5</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Females</td>
<td>12</td>
<td>70.6</td>
<td>23</td>
<td>57.5</td>
<td>32</td>
</tr>
<tr>
<td>Age Range (X=20)</td>
<td>17 - 20</td>
<td>14</td>
<td>82.4</td>
<td>25</td>
<td>62.5</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>21 - 49</td>
<td>3</td>
<td>17.6</td>
<td>15</td>
<td>37.5</td>
<td>17</td>
</tr>
<tr>
<td>Student Status</td>
<td>Undergrad</td>
<td>17</td>
<td>100.0</td>
<td>39</td>
<td>97.5</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>Graduate</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>2.5</td>
<td>5</td>
</tr>
<tr>
<td>Italian Major</td>
<td>Yes</td>
<td>1</td>
<td>5.9</td>
<td>4</td>
<td>10.0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>16</td>
<td>94.1</td>
<td>36</td>
<td>90.0</td>
<td>53</td>
</tr>
<tr>
<td>Speaks Dialect</td>
<td>Yes</td>
<td>1</td>
<td>5.9</td>
<td>1</td>
<td>2.5</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>16</td>
<td>94.1</td>
<td>39</td>
<td>97.5</td>
<td>46</td>
</tr>
<tr>
<td>Additional Language Study</td>
<td>Yes</td>
<td>9</td>
<td>52.9</td>
<td>11</td>
<td>27.5</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>8</td>
<td>47.1</td>
<td>29</td>
<td>72.5</td>
<td>37</td>
</tr>
</tbody>
</table>
### TABLE D.2

ONE-FACTOR MANOVAs ON SEVEN LANGUAGE MEASURES BY UNIVERSITY AND BY CLASS

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>df</th>
<th>ROOT</th>
<th>dfHYP</th>
<th>dfERR</th>
<th>F</th>
<th>p&lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>University</td>
<td>2</td>
<td>1</td>
<td>14.00</td>
<td>210.00</td>
<td>2.88</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(A)</td>
<td>2</td>
<td>6.00</td>
<td>105.50</td>
<td>2.62</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>S/A</td>
<td>111</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class (A)</td>
<td>7</td>
<td>1</td>
<td>49.00</td>
<td>512.11</td>
<td>2.73</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>36.00</td>
<td>475.49</td>
<td>1.49</td>
<td>.04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>25.00</td>
<td>428.09</td>
<td>1.22</td>
<td>.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>16.00</td>
<td>367.98</td>
<td>.90</td>
<td>.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td>5</td>
<td>9.00</td>
<td>293.70</td>
<td>.64</td>
<td>.76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6</td>
<td>4.00</td>
<td>205.00</td>
<td>.64</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7</td>
<td>1.00</td>
<td>103.00</td>
<td>.03</td>
<td>.88</td>
</tr>
<tr>
<td>S/A</td>
<td>106</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>113</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE D.3
UNIVARIATE ANALYSES OF VARIANCE FOR SEVEN LANGUAGE MEASURES BY UNIVERSITY AND BY CLASS

<table>
<thead>
<tr>
<th>CRITERION</th>
<th>MS</th>
<th>F (2,111)</th>
<th>P &lt;</th>
<th>MS</th>
<th>F (7,106)</th>
<th>P &lt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td>330.49</td>
<td>6.39</td>
<td>.002</td>
<td>337.68</td>
<td>8.86</td>
<td>.001</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>265.43</td>
<td>6.16</td>
<td>.003</td>
<td>286.94</td>
<td>9.20</td>
<td>.001</td>
</tr>
<tr>
<td>Cloze</td>
<td>63.13</td>
<td>4.45</td>
<td>.014</td>
<td>87.77</td>
<td>8.57</td>
<td>.001</td>
</tr>
<tr>
<td>Random</td>
<td>2105.76</td>
<td>9.25</td>
<td>.001</td>
<td>807.79</td>
<td>3.59</td>
<td>.002</td>
</tr>
<tr>
<td>Coherent</td>
<td>2868.56</td>
<td>7.96</td>
<td>.001</td>
<td>1757.75</td>
<td>5.58</td>
<td>.001</td>
</tr>
<tr>
<td>Recall</td>
<td>4.75</td>
<td>5.10</td>
<td>.008</td>
<td>3.65</td>
<td>4.42</td>
<td>.001</td>
</tr>
<tr>
<td>Recognition</td>
<td>3.66</td>
<td>3.84</td>
<td>.024</td>
<td>2.67</td>
<td>3.00</td>
<td>.006</td>
</tr>
</tbody>
</table>
### TABLE D.4

**DESCRIPTIVE STATISTICS OF SEVEN VERBAL MEASURES BY UNIVERSITY**

<table>
<thead>
<tr>
<th>Measure</th>
<th>University I</th>
<th>University II</th>
<th>University III</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Means (SDs)</td>
<td>Means (SDs)</td>
<td>Means (SDs)</td>
</tr>
<tr>
<td>Syntax</td>
<td>26.12 (5.73)</td>
<td>24.78 (6.44)</td>
<td>20.44 (8.01)</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>30.82 (5.19)</td>
<td>30.78 (5.15)</td>
<td>26.47 (7.70)</td>
</tr>
<tr>
<td>Cloze</td>
<td>8.12 (3.35)</td>
<td>7.03 (3.00)</td>
<td>5.37 (4.32)</td>
</tr>
<tr>
<td>Random</td>
<td>54.35 (14.63)</td>
<td>42.25 (16.89)</td>
<td>36.56 (13.84)</td>
</tr>
<tr>
<td>Coherent</td>
<td>65.88 (17.20)</td>
<td>54.90 (19.52)</td>
<td>45.90 (19.08)</td>
</tr>
<tr>
<td>Recall*</td>
<td>.60 (1.13)</td>
<td>.08 (.90)</td>
<td>-.23 (.96)</td>
</tr>
<tr>
<td>Recognition*</td>
<td>.61 (1.21)</td>
<td>-.11 (.87)</td>
<td>-.10 (.97)</td>
</tr>
</tbody>
</table>

*Statistics based on standardized scores.*
<table>
<thead>
<tr>
<th>MEASURE</th>
<th>CLASS</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Syntax</td>
<td>Mean</td>
<td>26.17</td>
<td>26.09</td>
<td>22.67</td>
<td>26.00</td>
<td>23.10</td>
<td>29.89</td>
<td>25.46</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>6.56</td>
<td>5.58</td>
<td>5.50</td>
<td>7.08</td>
<td>4.98</td>
<td>7.62</td>
<td>7.41</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Mean</td>
<td>28.00</td>
<td>32.36</td>
<td>31.50</td>
<td>30.96</td>
<td>29.90</td>
<td>34.22</td>
<td>32.27</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>5.44</td>
<td>4.57</td>
<td>4.55</td>
<td>5.52</td>
<td>4.93</td>
<td>5.40</td>
<td>5.20</td>
</tr>
<tr>
<td>Cloze</td>
<td>Mean</td>
<td>9.50</td>
<td>7.36</td>
<td>7.67</td>
<td>7.29</td>
<td>6.00</td>
<td>10.11</td>
<td>8.36</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>3.15</td>
<td>3.36</td>
<td>3.01</td>
<td>3.06</td>
<td>4.01</td>
<td>3.04</td>
<td>3.19</td>
</tr>
<tr>
<td>Random</td>
<td>Mean</td>
<td>54.17</td>
<td>54.46</td>
<td>38.17</td>
<td>45.75</td>
<td>36.30</td>
<td>40.22</td>
<td>42.09</td>
</tr>
<tr>
<td>Coherent</td>
<td>Mean</td>
<td>61.33</td>
<td>68.36</td>
<td>49.00</td>
<td>57.88</td>
<td>51.30</td>
<td>62.22</td>
<td>57.27</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>17.43</td>
<td>17.39</td>
<td>10.71</td>
<td>21.50</td>
<td>18.55</td>
<td>18.54</td>
<td>18.87</td>
</tr>
<tr>
<td>Recall*</td>
<td>Mean</td>
<td>0.51</td>
<td>0.65</td>
<td>-0.22</td>
<td>0.09</td>
<td>0.23</td>
<td>0.63</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.81</td>
<td>0.62</td>
<td>0.69</td>
<td>0.96</td>
<td>0.87</td>
<td>1.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Recognition*</td>
<td>Mean</td>
<td>0.21</td>
<td>0.82</td>
<td>-0.43</td>
<td>-0.07</td>
<td>-0.03</td>
<td>0.80</td>
<td>-0.09</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>1.56</td>
<td>1.00</td>
<td>1.18</td>
<td>0.77</td>
<td>0.95</td>
<td>0.81</td>
<td>0.91</td>
</tr>
</tbody>
</table>

* Statistics based on standardized scores.
<table>
<thead>
<tr>
<th>SOURCE</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Subjects</td>
<td>113</td>
<td>8943.90</td>
<td>4471.95</td>
<td>11.15**</td>
</tr>
<tr>
<td>A (University)</td>
<td>2</td>
<td>8943.90</td>
<td>4471.95</td>
<td>11.15**</td>
</tr>
<tr>
<td>B (Context Clueing)</td>
<td>1</td>
<td>160.29</td>
<td>160.29</td>
<td>.40</td>
</tr>
<tr>
<td>C (Boundary Markers)</td>
<td>1</td>
<td>483.99</td>
<td>483.99</td>
<td>1.21</td>
</tr>
<tr>
<td>D (Sequence)</td>
<td>1</td>
<td>118.84</td>
<td>118.84</td>
<td>.30</td>
</tr>
<tr>
<td>AB</td>
<td>2</td>
<td>3136.17</td>
<td>1568.09</td>
<td>3.91*</td>
</tr>
<tr>
<td>AC</td>
<td>2</td>
<td>1077.61</td>
<td>538.81</td>
<td>1.34</td>
</tr>
<tr>
<td>AD</td>
<td>2</td>
<td>361.07</td>
<td>180.53</td>
<td>.45</td>
</tr>
<tr>
<td>BC</td>
<td>1</td>
<td>6.75</td>
<td>6.75</td>
<td>.02</td>
</tr>
<tr>
<td>BD</td>
<td>1</td>
<td>465.59</td>
<td>465.59</td>
<td>1.61</td>
</tr>
<tr>
<td>CD</td>
<td>1</td>
<td>18.59</td>
<td>18.59</td>
<td>.05</td>
</tr>
<tr>
<td>ABC</td>
<td>2</td>
<td>186.43</td>
<td>93.22</td>
<td>.23</td>
</tr>
<tr>
<td>ABD</td>
<td>2</td>
<td>293.59</td>
<td>146.79</td>
<td>.37</td>
</tr>
<tr>
<td>ACB</td>
<td>2</td>
<td>851.14</td>
<td>425.57</td>
<td>1.06</td>
</tr>
<tr>
<td>ABD</td>
<td>2</td>
<td>293.59</td>
<td>146.79</td>
<td>.37</td>
</tr>
<tr>
<td>ADC</td>
<td>2</td>
<td>851.14</td>
<td>425.57</td>
<td>1.06</td>
</tr>
<tr>
<td>BCD</td>
<td>1</td>
<td>217.89</td>
<td>217.89</td>
<td>.54</td>
</tr>
<tr>
<td>ABCD</td>
<td>2</td>
<td>345.37</td>
<td>172.68</td>
<td>.43</td>
</tr>
<tr>
<td>S/ABCD</td>
<td>90</td>
<td>36092.86</td>
<td>401.03</td>
<td></td>
</tr>
</tbody>
</table>
Table D.6 continued

<table>
<thead>
<tr>
<th>Within Subjects</th>
<th>114</th>
<th>5160.01</th>
<th>5160.01</th>
<th>55.17***</th>
</tr>
</thead>
<tbody>
<tr>
<td>E (Task)</td>
<td>1</td>
<td>5160.01</td>
<td>5160.01</td>
<td>55.17***</td>
</tr>
<tr>
<td>AE</td>
<td>2</td>
<td>103.21</td>
<td>51.60</td>
<td>.55</td>
</tr>
<tr>
<td>BE</td>
<td>1</td>
<td>28.33</td>
<td>28.33</td>
<td>.30</td>
</tr>
<tr>
<td>CE</td>
<td>1</td>
<td>.23</td>
<td>.23</td>
<td>.00</td>
</tr>
<tr>
<td>DE</td>
<td>1</td>
<td>6617.70</td>
<td>6617.70</td>
<td>70.75***</td>
</tr>
<tr>
<td>ABE</td>
<td>2</td>
<td>21.96</td>
<td>10.98</td>
<td>.12</td>
</tr>
<tr>
<td>ACE</td>
<td>2</td>
<td>123.87</td>
<td>61.94</td>
<td>.66</td>
</tr>
<tr>
<td>ADE</td>
<td>2</td>
<td>56.65</td>
<td>26.32</td>
<td>.28</td>
</tr>
<tr>
<td>BCE</td>
<td>1</td>
<td>436.03</td>
<td>436.03</td>
<td>4.66*</td>
</tr>
<tr>
<td>BDE</td>
<td>1</td>
<td>2.00</td>
<td>2.00</td>
<td>.02</td>
</tr>
<tr>
<td>CDE</td>
<td>1</td>
<td>118.52</td>
<td>118.52</td>
<td>1.27</td>
</tr>
<tr>
<td>ABCE</td>
<td>2</td>
<td>1446.79</td>
<td>223.40</td>
<td>2.39</td>
</tr>
<tr>
<td>ABDE</td>
<td>2</td>
<td>219.96</td>
<td>109.98</td>
<td>1.18</td>
</tr>
<tr>
<td>ACDE</td>
<td>2</td>
<td>81.25</td>
<td>40.63</td>
<td>.43</td>
</tr>
<tr>
<td>BCDE</td>
<td>1</td>
<td>40.79</td>
<td>40.79</td>
<td>.44</td>
</tr>
<tr>
<td>ABCDE</td>
<td>2</td>
<td>183.71</td>
<td>91.85</td>
<td>.98</td>
</tr>
<tr>
<td>SE/ABCD</td>
<td>90</td>
<td>8416.98</td>
<td>93.52</td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>227</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** p < .0001
** p < .0005
* p < .05
<table>
<thead>
<tr>
<th>Clue</th>
<th>University</th>
<th></th>
<th></th>
<th>Level Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>II</td>
<td>III</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>57.00</td>
<td>43.62</td>
<td>46.15</td>
<td>48.93</td>
</tr>
<tr>
<td></td>
<td>(13.17)</td>
<td>(14.45)</td>
<td>(16.22)</td>
<td>(15.63)</td>
</tr>
<tr>
<td>No</td>
<td>61.46</td>
<td>54.76</td>
<td>36.45</td>
<td>50.89</td>
</tr>
<tr>
<td></td>
<td>(12.81)</td>
<td>(14.57)</td>
<td>(10.50)</td>
<td>(16.06)</td>
</tr>
<tr>
<td>Level Totals</td>
<td>59.23</td>
<td>49.19</td>
<td>41.30</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

DESCRIPTION OF CANONICAL ANALYSIS
Overview of Canonical Variate Analysis

In many respects, canonical variate analysis subsumes all other parametric techniques, however, it is most often explained as the multivariate extension of multiple regression analysis. Although the computational procedures are quite complex, the method of canonical analysis is straightforward.

The canonical correlation approach is most valuable when seeking information about the underlying relations between two sets of measures. The two sets \((X, Y)\) may, but need not be composed of independent (predictor) and dependent (criterion) variables respectively. For the sake of discussion, the \(X\) set is often referred to as the left set, and \(Y\) as the right set.

Given the two sets of measures gathered on the same group of subjects, matrices can be obtained that reflect the correlations among elements within each set \((R_{XX}, R_{YY})\), as well as the correlations of variables between sets \((R_{XY}, R_{YX})\). The object of canonical analysis is to find the strongest link between sets. It begins by partitioning the correlation matrix as described, then compounding the variables in the \(X\) and \(Y\) sets to produce two canonical variates that reflect the variance within the variable sets. To accomplish this, canonical selects weights or values for the original variables in each set such that when the weighted variables contained in a set are summed or packaged together, the combination in the right set is maximally related to the combination in the left. In more technical terms, canonical forms a pair of linear composites of the original variables in each set and solves for the weights such that the correlation between the canonical variates is the maximum possible.
Having found the correlation between the first pair of factors, canonical will continue to form a new pair of factors and repeat the process of solving for the weights to produce another canonical correlation orthogonal to the first. The number of solutions canonical can generate is equal to the number of original variables in the smaller of the two sets. In the present study, for example, the Y set was composed of only two measures (recall and recognition), thus allowing for only two canonical functions or roots.

As with other correlational methods, tests can be applied to determine the statistical significance of the canonical correlations obtained for each pair of variates generated. In canonical, the Wilks' Lambda ($\lambda$) associated with each root is converted to a chi-square or $F$ distribution. Most problems, however, require an explanation of more substance with respect to the nature and magnitude of the association than either the canonical weights or the canonical correlation squared can provide.

Information concerning the contribution of each element in a set to the calculation of the new variate is contained in the standardized weights yielded by the canonical correlation. These weights, however, are unreliable indicators of the composition or content of the new variate, since very often, intercorrelations among the variables in a set can produce false impressions of a particular variable's importance (see Darlington, 1968). In overcoming these problems, one can interpret instead, the structure coefficients associated with each variate that represent the correlations between the canonical variate scores and the original data for the variables within a set. The structure coefficients
permit one to understand the content of each pair of canonical variates much in the same way factors in factor analysis are interpreted via factor loadings.

Having grasped some notion of the content manifested by each new measure, the next step is to assess the strength or magnitude of the relationships between these measures. There are different ways of approaching this problem, but interpreting the canonical correlation or the square of that coefficient are not among them. By definition, the canonical correlation is the maximum linkage between sets, and is, therefore, an exaggerated index of the strength of the relationship. Similarly, squaring the canonical correlation results in an overestimate of the amount of variance shared by the two sets, because the number of canonical correlations is greater than one. A measure somewhat analogous to an average of the squared canonical correlation is warranted to meaningfully assess the magnitude of the relationship. One such measure, quotes "redundancy", formulated by Stewart and Love (1968), functions as does the $R^2$ in multiple regression. This measure specifies the amount of variance contained within a variable set that is absorbed by a given canonical variate for that set, and is computed by squaring the structure coefficients for each individual variable and summing those values. The amount of variance common to both sets, or the amount of information in one set that can be explained given the information in the other set is represented by the redundancy. The greater the redundancy, or overlap between sets, the more explanatory power. Hence, redundancy provides a meaningful assessment of the magnitude of the association between sets composed of multiple variables.
More detailed explanations of canonical variate analysis, including guidelines for interpreting the canonical factors and their relationships with each other, are presented in Cooley and Lohnes (1971), Levine (1977), Alpert and Peterson (1972), and Kerlinger (1973).