INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
THE ROLE OF LANGUAGE PROFICIENCY, TEST ANXIETY, AND TESTING PREFERENCES IN ESL STUDENTS' TEST PERFORMANCE IN CONTENT-AREA COURSES

DISSERTATION

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

By

Annela Teemant, B.A., M.A.

* * * *

The Ohio State University

1997

Dissertation Committee:
Professor Keiko Samimy, Adviser
Professor Charles R. Hancock
Professor John J. Kennedy

Approved by
Professor Keiko Samimy
Adviser
College of Education
ABSTRACT

This study investigates the jeopardy in content-area testing for English as a Second Language (ESL) students in university settings. A theoretical model is proposed which suggests that language proficiency, test anxiety, and test preferences are variables which predict ESL students' performance on content-area tests. Language proficiency was measured using the Michigan Test of English Proficiency, and a writing assessment. Test anxiety was comprised of both state- and trait-anxiety measures. Testing preferences assessed students' attitudes toward true/false, multiple-choice, short-answer/completion, restricted-response, and extended-essay formats using a semantic differential scale.

Both qualitative and quantitative data were gathered. First, thirteen ESL students were interviewed, and a content analysis of the data provides a general overview of the problems, preferences, strategies, and teacher practices related to content-area testing for ESL students. Second, quantitative data relating to language, anxiety, preferences, and test performance were gathered in a specific course—American Heritage—with 38 ESL students at Brigham Young University. Data were analyzed using a multiple regression approach investigating the relationships among variables in the theoretical model (i.e., path analysis). In addition, four American Heritage ESL students were interviewed.

The qualitative and quantitative findings suggest that the testing jeopardy for ESL students is a multidimensional phenomenon, with aspects internal and external to the student's control. In the course-specific context, the theoretical model accounted for 60% of the variance in test performance. Language proficiency and testing preferences did not account for a significant amount of variance; however, test anxiety was found to be the
strongest predictor of students' content-area test performance. This study provides qualitative and quantitative evidence that the American Heritage professor's instructional and testing accommodations minimized the language proficiency and testing preference jeopardy in testing for ESL students. Implications for theory, pedagogy, and research are presented.
Dedicated to my father Mihkel Teemant
ACKNOWLEDGMENTS

Completion of this dissertation is the result of many patient faculty members, gracious students, and long-suffering family and friends. At The Ohio State University, I wish to thank Elizabeth Bernhardt for being an insightful and demanding advisor (1993-95). I am also indebted to Keiko Samimy for serving willingly and with empathy as my generals and dissertation advisor (1995-1997). Charles Hancock, Gilbert Jarvis, and John Kennedy have also made substantial contributions to my graduate studies. My thanks to Kathy Romstedt and her ESL colleagues at the American Language Program for scoring writing assessments as well.

At Snow College (Ephraim, Utah), I wish to thank Professors Kerry Hansen, Mary Ann Christison, and Gerhard Bolli, and the staff and students who were involved in my pilot study.

I am deeply indebted to the generous enthusiasm of many colleagues at Brigham Young University (Provo, Utah). Professor William Fox and his teaching assistant Kimo Pokini cooperation exceeded all my expectations. My research depended on their consent, their collaboration, and their students. In addition, I am grateful to Professors Cheryl Brown (my M.A. advisor), Mary Lee Scott, and Diane Strong-Krause who continue to be friends and models of professionalism.

I gratefully acknowledge the emotional, and intellectual support of cherished friends: Cami Courtright, Ann Rozyczki, Sue Granoff, Marinella Garatti, Macel Braxton, Anna Gelinas, Teresa Benedetti, Nancy Giraldo, and Rebecca Brennan. As always, I humbly acknowledge the power of my mother, Aino Kivimaa Teemant, and my sister, Kaare Teemant, to motivate me to strive and endure.
VITA

October 25, 1961 .......................... Born - San Mateo, California

1983 ............................................... B.A., Brigham Young University, Provo, Utah

1987 .................................................. TESL Certificate, Brigham Young University

1988 ............................................... M.A., Brigham Young University

1988 ............................................... ESL/ESP Instructor, University of Jyväskylä, Jyväskylä, Finland

1989-1991 ........................................ ESL Instructor/Supervisor, Brigham Young University


1993-present .....................................Graduate Teaching and Research Associate, The Ohio State University

PUBLICATIONS

Research Publication


FIELDS OF STUDY

Major Field: Education

Minor Field: Quantitative Statistics
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>VITA</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
<tr>
<td><strong>CHAPTER 1: THE PROBLEM</strong></td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Significance of the Problem</td>
<td>4</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>8</td>
</tr>
<tr>
<td>Research Questions</td>
<td>10</td>
</tr>
<tr>
<td>Basic Assumptions</td>
<td>12</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>13</td>
</tr>
<tr>
<td>Delimitations and Limitations of the Study</td>
<td>16</td>
</tr>
<tr>
<td>Chapter Summary</td>
<td>17</td>
</tr>
<tr>
<td><strong>CHAPTER 2: REVIEW OF LITERATURE</strong></td>
<td>18</td>
</tr>
<tr>
<td>Language Proficiency</td>
<td>18</td>
</tr>
<tr>
<td>Test Anxiety</td>
<td>23</td>
</tr>
<tr>
<td>Testing Preferences</td>
<td>28</td>
</tr>
<tr>
<td>ESL/LM Test Performance</td>
<td>31</td>
</tr>
<tr>
<td>Predicting Academic Success</td>
<td>34</td>
</tr>
<tr>
<td>Theoretical Perspective</td>
<td>38</td>
</tr>
<tr>
<td>Overall Conclusions</td>
<td>40</td>
</tr>
<tr>
<td><strong>CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY</strong></td>
<td>43</td>
</tr>
<tr>
<td>Introduction</td>
<td>43</td>
</tr>
<tr>
<td>Nonexperimental Research</td>
<td>44</td>
</tr>
<tr>
<td>Qualitative Research</td>
<td>45</td>
</tr>
<tr>
<td>Participants</td>
<td>46</td>
</tr>
<tr>
<td>Qualitative Interview Subjects</td>
<td>46</td>
</tr>
<tr>
<td>Course Subjects</td>
<td>48</td>
</tr>
</tbody>
</table>
### CHAPTER 4: QUALITATIVE INTERVIEW RESULTS AND DISCUSSION

**Introduction**

**Results**

- Problems in Testing
  - 1. The Context
  - 2. The Testing Culture
  - 3. The Language
- Testing Preferences
- Testing Strategies
- Helpful Teacher Practices

**Discussion**

- 1. Language Proficiency
- 2. Test Anxiety
- 3. Testing Preferences
- General Implications
- Limitations
- Future Directions
- Conclusion

### CHAPTER 5: COURSE-SPECIFIC RESULTS AND DISCUSSION

**Introduction**

**Qualitative Results**

**The Interviews**

- 1. Problems in Testing
  - A. The Context
  - B. The Testing Culture
  - C. Other Concerns
- 2. Testing Preferences
- 3. Testing Strategies
- 4. Helpful Teacher Practices

**Course Evaluations**

**Qualitative Discussion**

**Conclusion**

**Quantitative Results**

- Test Performance
- Language Proficiency
- Test Anxiety
  - 1. Trait-Anxiety
  - 2. State-Anxiety
- Testing Preferences
## Theoretical Model
- Stage One: 121
- Stage Two: 127
- Stage Three: 128

### Discussion of Findings
- 1. Language Proficiency: 132
- 2. Test Anxiety: 134
- 3. Testing Preferences: 135

### General Implications
- 1. Language Proficiency: 138
- 2. Test Anxiety: 139
- 3. Testing Preferences: 139

### Future Directions
- Conclusion: 140

### Conclusion
- 1. Language Proficiency: 141
- 2. Test Anxiety: 141
- 3. Testing Preferences: 141

## CHAPTER 6: CONCLUSIONS
- Introduction: 142

### Synthesis of Findings
- 1. The Quantitative Profile: Problems and Practices: 142
- 2. Language Proficiency: 143
- 3. Test Anxiety: 146
- 4. Testing Preferences: 146

### Synthesis of Implications
- 1. Theoretical Implications: 147
- 2. Classroom Implications:
  - A. ESL Programs: 148
  - B. Teacher Preparation Programs: 149
  - C. Content-Area Collaboration: 149
- 3. Research Implications: 150

### Limitations
- 1. Language Proficiency: 150
- 2. Test Anxiety: 150
- 3. Testing Preferences: 150

### Future Directions
- Conclusion: 151

### APPENDICES
- A. Relational Learning Process: 152
- B. Semi-Structured Interview Questions: 154
- C. Examples of Language Proficiency Test Items: 156
- D. Example Writing Prompts and Rating Criteria: 158
- E. State-Anxiety Measure: 160
- F. Testing Preference Questionnaire: 167
- G. Sample Test Questions from American Heritage 100: 175
- H. Example Transcript: Subject 08: 180
- I. Test Format Examples: 198
- J. Reasons for Attitudes Toward Test Formats: 200

### REFERENCES
- 205
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>Description of General Interview Subjects</td>
<td>47</td>
</tr>
<tr>
<td>3.2</td>
<td>Description of Course-Specific Interview Subjects</td>
<td>49</td>
</tr>
<tr>
<td>4.1</td>
<td>ESL/LM student strategies for coping with content-area tests</td>
<td>77</td>
</tr>
<tr>
<td>4.2</td>
<td>ESL/LM students' requests of content-area teachers</td>
<td>79</td>
</tr>
<tr>
<td>5.1</td>
<td>Mean rating of how effectively course components reinforced relational learning</td>
<td>96</td>
</tr>
<tr>
<td>5.2</td>
<td>Mean scores of students' test performance by format</td>
<td>102</td>
</tr>
<tr>
<td>5.3</td>
<td>Test format preferred for best demonstrating content knowledge on tests</td>
<td>103</td>
</tr>
<tr>
<td>5.4</td>
<td>Mean language proficiency scores by skill area</td>
<td>104</td>
</tr>
<tr>
<td>5.5</td>
<td>Students' perceptions of their strongest language skill</td>
<td>105</td>
</tr>
<tr>
<td>5.6</td>
<td>The language skill that students feel helps most when taking tests</td>
<td>105</td>
</tr>
<tr>
<td>5.7</td>
<td>Summary statistics for the debilitative and facilitative scales of trait-anxiety</td>
<td>107</td>
</tr>
<tr>
<td>5.8</td>
<td>Mean score state-anxiety ratings on course exams by question format</td>
<td>110</td>
</tr>
<tr>
<td>5.9</td>
<td>Question formats student feel best allow them to demonstrate knowledge</td>
<td>112</td>
</tr>
<tr>
<td>5.10</td>
<td>Mean composite ratings of question formats on a semantic differential scale</td>
<td>113</td>
</tr>
<tr>
<td>5.11</td>
<td>Mean score ratings of test formats by bipolar adjective pairs</td>
<td>114</td>
</tr>
<tr>
<td>5.12</td>
<td>Mean scores for production and recognition formats by adjective pairs</td>
<td>117</td>
</tr>
<tr>
<td>5.13</td>
<td>Multiple regression analysis of trait-anxiety on language group, and gender</td>
<td>121</td>
</tr>
</tbody>
</table>
Table | Page
--- | ---
5.14 Multiple regression analysis of testing preferences on language group, and gender | 122
5.15 Multiple regression analysis of production format preferences on language group, and gender | 124
5.16 Multiple regression analysis of test performance on testing preferences, trait-anxiety, language group, and gender | 125
5.17 Means, standard deviations, and correlations for testing performance, language proficiency, testing preferences, trait-anxiety, language group, and gender | 128
5.18 Multiple regression analysis of content-area test performance on language proficiency, testing preferences, trait-anxiety, language group, and gender | 129
5.19 Multiple regression analysis of full model showing change in $R^2$ | 131
6.1 General interview and American Heritage profiles of content-area testing | 144
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>A visual model of the variables which are potentially important in predicting ESL/LM students' test performance in content-area classes</td>
<td>41</td>
</tr>
<tr>
<td>3.1</td>
<td>The data collection process</td>
<td>58</td>
</tr>
<tr>
<td>3.2</td>
<td>A path diagram of the variables to be investigated in predicting ESL/LM students' test performance in content-area classes</td>
<td>63</td>
</tr>
<tr>
<td>5.1</td>
<td>Stage one path diagram: Influence of language group and gender on trait-anxiety</td>
<td>122</td>
</tr>
<tr>
<td>5.2</td>
<td>Stage one path diagram: Influence of language group and gender on testing preferences</td>
<td>124</td>
</tr>
<tr>
<td>5.3</td>
<td>Stage one path diagram: Influence of testing preferences, trait-anxiety, language group, and gender on test performance</td>
<td>126</td>
</tr>
<tr>
<td>5.4</td>
<td>Stage two path diagram: Influence of testing preferences, trait-anxiety, language group, and gender on language proficiency</td>
<td>127</td>
</tr>
<tr>
<td>5.5</td>
<td>Stage three path diagram: Influence of language proficiency, testing preferences, trait-anxiety, language group, and gender on content-area test performance in American Heritage 100</td>
<td>130</td>
</tr>
</tbody>
</table>
CHAPTER 1

THE PROBLEM

Introduction

For United States educators, two themes dominate the end of the twentieth century, and will lead educators into the twenty-first century. The first theme entails the changing demographics of the United States, and its impact on schooling. The second theme focuses on the national education reform movement, which is designed to raise educational standards to a "world class level," (Stansfield, 1994). The purpose of this chapter is to outline how these current educational themes connect the issues of language learning and content learning, and justify investigation of content-area testing from a second language perspective.

Over the past decade, there has been a dramatic increase in the number of English-as-a-second-language (ESL) students, from both immigrant and international backgrounds, entering American schools—primary and secondary schools as well as colleges and universities (Rosenthal, 1996; Steward, 1991). Described as "rapidly growing." (Rosenthal, 1992/1993, p. 185), "escalating exponentially" (Faltis and Hudelson, 1994, p. 460), and representing a "burgeoning level of diversity," (Short, 1994, p. 581), language-minority (LM) students will comprise 42% of the public school population (Carrasquillo & Rodriguez, 1995), or approximately 3.4 million students by the turn of the century (Lam, 1993). More than six million children in the United States do not use English as their native language at home (Rosenthal, 1996).
At the tertiary level, the student population is also diversifying. In 1992 *The Chronicle of Higher Education Almanac* reported a 54% increase between 1980 and 1990 of students from Asian, Hispanic, American Indian, and foreign backgrounds in the undergraduate population—up from 978,000 to 1,508,000 (cited in Rosenthal, 1996, p. 20).

This linguistic and cultural diversity has prompted concern among educators, in all disciplines and in all school settings, over how to promote and support the academic success of these learners. Equity is a pivotal aspect of this new era of diversity. While much of the literature focuses on addressing the needs of ESL/LM students in primary and secondary schools, these same realities need to be addressed in higher education (Snow & Briton, 1988; Steward, 1991). Grant (1992) asserts that all schools "are being required to respond to the challenge of... making schooling equal and equitable for all," (p. 1).

For LaCelle-Peterson and Rivera (1994), "educational excellence" is described as being "designed with all learners in mind—including those who bring linguistic riches with them to school," (p. 73). However, they also note that, historically, any advances in equity made on behalf of ESL/LM students has generally been the result of "litigation and legislation" not the result of a "disinterested pursuit of educational excellence," (p. 57).

The *Lau* decision, for example, focused national attention on "educational equity for LM students," (Lucas, Henze, & Donato, 1990, p. 316). Grant and Millar (1992) explain what is meant by 'equity', noting that "'equal opportunity', meaning having equal access is not synonymous with 'equity', which means having a fair and just opportunity," (p. 14). The *Lau* decision clarified that being placed in an English-medium classroom satisfied the requirement for access, but did not provide "equality of outcome" if students were not also provided with the necessary English language background to succeed (McKay & Freedman, 1990, p. 399).

The U.S. educational system has entered a "new era" that demands educators to re-evaluate current practices, and demonstrate a "readiness -- and a willingness -- to adapt" to
the challenges and opportunities of the twenty-first century, with its inevitable diversity (Steward, 1991, p. 25). Currently, most content-area teachers, regardless of their field or school setting, remain ill-prepared, in terms of pedagogy and assessment, to work with or accommodate diverse student populations (Clair, 1995; Gollnick, 1992; Penfield, 1987; Rosenthal, 1992/1993; Simms & Leonard, 1986; Teemant, 1996). Preparing teachers to respond to the needs of ESL/LM students has become a "serious educational concern," (Carrasquillo & Rodriguez, 1995, p. 6).

Concurrent with the diversification of the U.S. student population is a national focus on educational reform, which began with the publication of A Nation at Risk in 1983, and the subsequent development of the National Educational Goals and Goals 2000: Educate America Act (H.R. 1804) (Lam, 1993; Stansfield, 1994). Short (1993) notes that "assessment dominates the educational reform dialogue," (p. 630). Testing is the tool being used by educators and politicians to "to reshape teaching and to effect learning in the schools," (Stansfield, 1994, p. 43).

Because multiculturalism is a reality in the classroom (Levy, 1993), the aims and outcomes of the reform movement, with its attention on testing, must be sensitive to and appropriate for all learners, including ESL/LM students. Bernhardt, Destino, Kamil, and Rodriguez-Muñoz (1995) claim that ESL/LM students "are in double jeopardy when confronted with assessment of any type," because they are "forced into demonstrating knowledge in a language over which they have only partial . . . control." (p. 6). Distinguishing between language knowledge and content knowledge is essential when testing ESL/LM students (Short, 1993; Brinton et al., 1989; Rosenthal, 1996).

Substantial research is needed to more fully understand testing from a second language perspective. Madaus (1994) argues for proponents of reform to consider "which individuals and groups are hurt by testing," (p. 81). Lemke (1990, p. 80) charges that present testing practices undeniably "favor" students from middle-class, standard dialect,
individualistic, and Protestant backgrounds. Anderson (1988) and Soto (1992) contend that testing is unfair when ESL/LM students are measured against a solely white, eurocentric, monolingual, normed yardstick, which yields "neither equitable nor valid results," (LaCelle-Peterson & Rivera, 1994, p. 66). According to Alexander and Parsons (1991), current testing practices continue to be a "highly ethnocentric" phenomenon, which "is particularly distant to the experiences and goals of minority populations," (p. 245).

In summary, changing demographics in the U.S. have clearly had an impact on schooling. Language learning issues have become the concern not only of ESL instructors but of every teacher serving ESL/LM students (Carrasquillo & Rodriguez, 1995). Furthermore, educators face serious obstacles in using testing to reform education. To date, what constitutes equitable content-area testing for ESL/LM students is undefined at the national, state, and classroom level. Little empirical research has focused on how language proficiency impacts content-area testing for ESL/LM students. Although language educators have traditionally only been concerned with assessing language ability, the current educational milieu affirms that content-area testing has emerged as a second language issue, worthy of systematic investigation.

Significance of the Problem

ESL educators have been at the forefront in advocating collaboration across disciplines to increase all teachers' awareness of the linguistic and cultural challenges ESL/LM students face in learning and demonstrating content knowledge through a second language (Benesch, 1988; Geisinger & Carlson, 1992; Harklau, 1994; Rosenthal, 1996; Teemant, 1996; Teemant, Bernhardt, & Rodriguez-Muñoz, 1996). Specifically, LaCelle-Peterson and Rivera (1994, p. 73) describe "new partnerships," Faltis and Hudelson (1994, p. 466) describe "new dialogue," and Hernández (1992, p. 143) describes the need for a "multidisciplinary" approach to understanding ESL/LM students' performance in content-area courses.
Hernández (1992) suggests that the research agenda surrounding ESL/LM students in content-area classes should focus on three factors: content, methodology, and testing; however, among these factors, content-area testing has received the least attention, even in transitional approaches specifically designed by ESL faculty (Brinton et al., 1989).

Knowing who, what, when, and how best to test ESL/LM students in content-area courses remains, from both an empirical and descriptive perspective, nebulous, underspecified, and perplexing in that language knowledge and content knowledge merge in the act of test-taking.

A student's English language proficiency is one factor influencing academic success; however, distinguishing between language issues and content issues in testing is complex (Brinton et al., 1989; Carrasquillo & Rodriguez, 1995; Hernández, 1992; Mohan, 1986; Rosenthal, 1996). LaCelle-Peterson and Rivera (1994) observe that, "every test will inevitably measure both what the learner knows about the particular subject matter and the learner's proficiency in the particular language," (p. 67). Short (1993) explains this complexity as follows:

because language and content are intricately intertwined, it is difficult to isolate one feature from the other in the assessment process. Thus, teachers may not be sure whether a student is simply unable to demonstrate knowledge because of a language barrier or whether, indeed, the student does not know the content material being assessed. Yet a distinction needs to be drawn, especially if a student is not succeeding in a course, (p. 629).

There is a substantial body of research aimed at predicting ESL/LM students' academic success in tertiary institutions. Graham (1987) reviewed 18 prediction studies, many of which relied on grade point average (GPA) as a measure of success; however, GPA is broadly considered an inexact measure. In her synthesis of these studies, Graham (1987) found that "the relationship between English language proficiency and academic success is murky," (p. 506). She concluded that language proficiency be seen as one factor, among many, influencing academic success. An ESL/LM student's gender, language background,
length of residency, university experience, or background in a particular content-area are other test-taker characteristics which may also influence test performance.

Few studies have focused on the role of language proficiency in success on specific national, state, or classroom examinations. Bernhardt et al. (1995) warn that there is danger in ignoring "the special challenges created by the addition of linguistic" factors when interpreting testing performance. Investigating test performance and the impact of language proficiency from a classroom perspective offers a potentially valuable alternative to the ambiguity characteristic of prediction studies focusing on GPA.

Another factor influencing test performance is test anxiety. Noting that U.S. culture is "test-conscious," Spielberger and Vagg (1995b) assert that test anxiety "has become a pervasive contemporary problem," (p. 3). Research has shown that test anxiety affects minority students to a greater degree than majority students (Phillips, Pitcher, Worsham, & Miller, 1980). Consequently, ESL/LM students who are ill-prepared for a testing situation may reduce their "access to educational and occupational opportunities" by testing poorly (Spielberger & Vagg, 1995a, p. xiii). While there is consensus regarding test anxiety's impact on test performance (e.g., Sapp, 1993; Spielberger & Vagg, 1995b), researchers have not investigated ESL/LM students' test anxiety in combination with other test-taker factors, such as language proficiency.

It is widely acknowledged that ESL/LM students struggle to represent accurately what they know on content-area tests (e.g., Rosenthal, 1996; Carrasquillo & Rodriguez, 1995; Brinton et al., 1989). According to LaCelle-Peterson and Rivera (1994, p. 69), current testing practices fail to "capture" what ESL/LM students really know in terms of content knowledge. For example, which test format (e.g., true/false, multiple-choice, short-answer/completion, restricted-response, and extended-essay questions) allows ESL/LM students to most accurately and completely demonstrate their content knowledge?
Unfortunately, ESL/LM students are often described only in terms of linguistic deficits; that is, their non-traditional, inadequate, disadvantaged, at-risk status, or their limited ability to compete with their language-majority peers in academic contexts (Anderson, 1988; Simms and Leonard, 1986; Soto, 1992). These linguistic concerns often overshadow the education, knowledge, skills, and abilities these students possess in their native or first languages (Rosenthal, 1996).

On a national level, Madaus (1994) maintains that "the high-stakes multiple-choice testing programs of the 1970s and 1980s did not measure the many important cognitive skills now called for by reformers," (p. 83). Educators admit standardized exams do not provide a complete record of any student's competence (Alexander & Parsons, 1991; Short, 1993); however, these exams have not only produced "invalid results" for ESL/LM students, but had particularly "negative accumulative influences on students' motivation, self-esteem, and attitude," (Lam, 1993, p. 180).

Nield and Wintre (1986) aptly note that anytime students are "indignant, anxious, or frustrated in the test situation... the validity of the test is in question," (p. 196). Yet when ESL/LM students do not succeed in various content-area courses, whether at the primary, secondary, or tertiary level, the blame is usually assigned to the student not the test (Lucas et al., 1990; Rosenthal, 1996; Soto, 1992). Poor test performance may lead teachers, administrators, or policy-makers to "search... for what is inherently 'wrong' with diverse learners," (Soto, 1992, p. 153), rather than pondering their own "inappropriate knowledge of how to respond to students who have difficulty," (Good, 1987, p. 33).

Because much of the literature in educational measurement and research addresses the reliability and validity of various examinations in differentiating, placing, or predicting success among students, the opinions, concerns, and preferences of test-takers themselves have often been ignored (Gellman & Berkowitz, 1993; Nield & Wintre, 1986; Zeidner,
Although research with native English speakers demonstrates that test-takers have clear preferences in testing, test developers and evaluators have not traditionally considered test-takers' needs a credible point of departure.

Perhaps, Madaus (1994) is correct in asserting that “we Americans cannot test, examine, or assess our way out of our educational problems,” (p 78). Addressing the testing concerns of ESL/LM students requires new ways of conceptualizing solutions. For example, rather than beginning with the test, researchers may gain important insights into testing for ESL/LM students by beginning with the test-takers themselves. Because the testing preferences of ESL/LM students have not been investigated extensively from descriptive or empirical perspectives, this avenue of research holds some promise in elucidating ESL/LM student performance on content-area tests.

In summary, ESL/LM students who are still in the process of acquiring English are not able to adequately represent their content knowledge on tests. ESL/LM students, whose abilities are thus underestimated on national, state, or classroom examinations, can suffer repercussions personally, educationally, or vocationally. Although language and content-area teachers, administrators, and policy-makers fervently discuss testing issues for ESL/LM students in theoretical terms, little research has been devoted to explicating testing from a second language perspective. Understanding ESL/LM students' success in content-area courses requires further investigation of how such test-taker factors as language proficiency, test anxiety, testing preferences, and personal characteristics (e.g., language group and gender) explain test performance in content-area classes.

**Purpose of the Study**

The purpose of this study is to investigate content-area testing from a second language perspective. Specifically, this study sets out to document qualitatively and quantitatively the experience ESL/LM students have in testing when required to express content knowledge in a second language. The relationship between test-taker factors
and test performance is investigated among 38 ESL/LM students in American Heritage courses at Brigham Young University (BYU) in Provo, Utah.

The study consists of four phases. Phase One is devoted to interviewing a minimum of 10 university-level ESL/LM students about their experience in taking content-area tests in their second language. This qualitative portion of the study will represent the insider's perspective on content-area testing. Furthermore, it will compliment and aid in the accurate interpretation of data gathered in the subsequent stages of the study.

Phase Two of the study entails gathering data via a questionnaire related to three test-taker variables: personal characteristics, testing preferences, and trait-anxiety (T-anxiety). Personal characteristics are defined as gender, language background (e.g., Spanish, Japanese, Arabic), length of residency, length of U.S. university experience, familiarity with U.S. examination system, and number of previous American history courses taken. Subjects' testing preferences will be defined using a semantic-differential, seven-point scale with ten bipolar adjective pairs (e.g., difficult-easy, tricky-straightforward, etc.). True/false, multiple-choice, short-answer/completion, restricted-response, and extended-essay questions will be considered. Trait-anxiety is defined in term of a student's general inclination toward test anxiety (Spielberger & Vagg, 1995b), as measured by the Alpert and Haber (1960) Achievement Anxiety Test (AAT).

Phase Three of the study consists of collecting various language and content-area test scores. For example, subjects' language proficiency in reading, writing, grammar, and vocabulary will be assessed using the Michigan Test of English Language Proficiency (MTELP), and an impromptu writing task. Over the course of a semester, subjects' scores on four American Heritage tests will also be recorded. Following each content-area test, subjects will complete a one-page state-anxiety (S-anxiety) questionnaire adapted from Madsen and Murray (1984). S-anxiety is defined as task-specific or situational anxiety.
(Spielberger & Vagg, 1995b). These data, combined with that gathered in Phase Two, will be used to delineate the relationship of the various test-taker variables to ESL/LM students' performance on content-area tests. In Phase Four, ten percent of the classroom subjects will be randomly selected for follow-up interviews regarding their testing experience in American Heritage 100.

These data will provide 1) a general overview of the testing preferences of ESL/LM students, independent of a particular course; and 2) a specific example of how language proficiency, test anxiety, testing preferences, and personal characteristics predict students' classroom test performance in a particular undergraduate, content-area course. The findings of this study will make a substantial contribution to educators' understanding of how to interpret test performance, and appropriately accommodate ESL/LM students in testing contexts.

Research Questions

This study investigates test performance in a content-area class from a second language perspective. The following research question guides the qualitative investigation of content-area testing:

1. What are ESL/LM students' opinions, concerns, and preferences related to being tested in content-area courses?

The following series of research questions guide the course-specific, quantitative investigation of how, and if, language proficiency, test anxiety, testing preferences, and the personal characteristics of gender and language group influence content-area test performance for ESL/LM students. Research questions related to endogenous variables (i.e., language proficiency, test anxiety, and testing preferences) are listed first, and followed by a research question which combines the exogenous variables of gender and language group with these endogenous variables to investigate a theoretically formulated
model of factors which may be important to understanding the second language experience in content-area testing for ESL/LM students.

**Language Proficiency**

2. Is there a relationship between language proficiency and test performance?

3. Which language skill or set of skills (i.e., reading, writing, grammar, or vocabulary) best predict/s overall performance on content-area tests?

4. Which language skill/s (i.e., reading, writing, grammar, or vocabulary) best predict/s performance on various test formats (i.e., multiple-choice, and restricted-response)?

**Test Anxiety**

5. Is there a relationship between trait-anxiety and test performance?

6. Is there a relationship between the personal characteristic of length of residency, length of U.S. university experience, familiarity with U.S. examination system, or number of previous American history courses taken and trait-anxiety?

7. Is there a relationship between state-anxiety and test performance?

8. Is there a relationship between the personal characteristic of length of residency, length of U.S. university experience, familiarity with U.S. examination system, or number of previous American history courses taken and state-anxiety?

**Testing Preferences**

9. What are ESL/LM students' testing preferences in content-area courses?

10. Is there a relationship between preference for each test format (e.g., restricted-response, and multiple-choice) and test performance on each test format?

11. Is there a relationship between the personal characteristic of length of residency, length of U.S. university experience, familiarity with U.S. examination system, or number of previous American history courses taken and testing preferences?
Theoretical Model

12. Are there relationships among gender, language group, trait-anxiety, testing preference, language proficiency, and test performance?

Basic Assumptions

In order to conduct this study, various assumptions were made in three areas: ESL/LM students, test-taker variables, and research processes. These assumptions follow.

ESL/LM Students:

1. ESL/LM students at the university setting are motivated to be academically successful.

2. No ESL program can completely prepare students for the knowledge, skills, and abilities needed to succeed in university content-area courses.

3. Because data will be collected from ESL/LM students at a private, religious university, students are assumed to be from a more advantaged than disadvantaged socio-economic background.

4. International ESL students and LM students (i.e., immigrants and refugees) do not always share the same economic, political, religious, or educational backgrounds.

5. A learner's language and cultural background play a role in academic success at U.S. universities.

Test-Taker Variables:

6. A student's English proficiency, which is a multidimensional construct, can be measured reliably using standardized language proficiency exams.

7. A student's familiarity, experience, and preferences regarding various testing formats are important in understanding test performance:

8. A student's trait- and state-anxiety can be measured reliably, and can affect test performance.
Research Processes:

9. Important insights into testing can be obtained by interviewing ESL/LM students, using qualitative methods, about their testing experiences in U.S. universities.

10. During semi-structured interviews, students will be able to answer questions honestly and candidly, revealing their opinions, concerns, and preferences regarding testing.

11. Quantitative analyses can provide useful empirical evidence, elucidating the relationships among various test-taker variables and test performance for ESL/LM students.

Definition of Terms

To clarify the use of terminology in this study, operational definitions of key independent variables are provided, followed by definitions of other relevant terminology, listed alphabetically.

Independent Variables:

Language proficiency is defined by scores on the MTELP, and an impromptu writing assignment. These scores indicate how well a learner can "read, write, speak, or understand language," (Richards, Platt, & Platt, 1992, p. 204).

Test anxiety is defined as a particular type of general anxiety related to fear of failing a test, and is measured using a variety of self-report inventories (Sapp, 1993). It is comprised of trait and state anxiety. For example, the Alpert and Haber (1960) Achievement Anxiety Test (AAT) combines facilitative and debilitative anxiety measures to yield a subject's trait anxiety score related to testing. Trait-anxiety is a person's stable, inclination toward anxiousness under stress (Spielberger, 1966). On the other hand, state-anxiety, indicated by scores on an adapted version of Madsen's and Murray's (1984) S-anxiety measure, is defined as a dynamic, transitory, task-specific emotional reaction to a test, which varies depending on such factors as the nature of the test questions, ability.
preparation, and individual personality differences (Sapp, 1993; Spielberger & Vagg, 1995b).

**Testing preferences**, as indicated by scores on an adapted version of Zeidner's (1987) questionnaire on format preferences, are defined by a student's semantic differential ratings of true/false, multiple-choice, short-answer, restricted-response, and extended-essay test formats.

**Other Terminology:**

**Achievement tests** measure what a student has learned with reference to a set of instructional learning outcomes (Gronlund, 1993). For the purposes of this study, achievement tests are limited to a specific course, and exclude national standardized achievement examinations.

**Content-area classes** is a term used among ESL professionals to describe subject-matter courses other than language. For example, science, math, history, or accounting classes are described as content-area classes.

**Endogenous variables** are defined as variables whose variation is caused by other variables (whether endogenous or exogenous) within a causal model (Kerlinger & Pedhazur, 1973; Vogt, 1993).

**English as a second language (ESL)** refers to the use of English by international, immigrant, or refugee groups in English-speaking countries (Richards, Platt, & Platt. 1992).

**Exogenous variables** are defined as variables whose variation is caused by variables outside the causal model/system being studied (Kerlinger & Pedhazur, 1973; Vogt, 1993).

**Extended-essay test items** allow students the most freedom in determining the "form and scope" of their responses to questions (Gronlund, 1993).

**Language-minority (LM)** is a term used to describe a group of people who have a language other than the dominant language of a particular country, for example Spanish,
Chinese, or Russian speakers in the U.S. (Richards, Platt, & Platt, 1992). This term emphasizes the ethnic as well as linguistic minority status of people. This group typically includes immigrants, refugees, or long-term residents who do not speak English in their homes, but does not include visiting international students. On the other hand, the term language-majority is used to describe native, English speakers in the U.S.

Limited English proficient (LEP) refers to a minority student in an English-speaking country (Richards, Platt, & Platt, 1992). The student's emerging proficiency in English does not allow full and/or successful participation in a classroom where English is the language of instruction.

Mainstreaming is the placement of ESL students into regular, content-area classes in a school where English is the language of instruction (Richards, Platt, & Platt, 1992).

Multiple-choice test items have a stem, and more than two alternatives for students to choose among as answers (Gronlund, 1993).

Native language (L1) or first language refers to a person's mother tongue, or the language they first acquired as children (Richards, Platt, & Platt, 1992).

Qualitative research is a process relying on words or non-numerical data, such as written documents, direct observation, or open-ended interview transcriptions, to investigate a social or human problem (Creswell, 1994).

Quantitative research is a process that tests if the generalizations of a theory are true; it uses variables and measures producing numerical data to investigate social or human problems from a statistical perspective (Creswell, 1994).

Restricted-response test items provide boundaries and limits to how a student answers a question; for example, a student will be asked to list, define or give reasons (Gronlund, 1993).

Second language (L2) is a term used to describe acquisition of any language other than a person's first or native language (Ellis, 1994).
Second language acquisition (SLA) is the process of learning and developing proficiency in a second or foreign language (Richards, Platt, & Platt, 1992).

Short-answer test items require a student to provide only a word, number, or symbol that answers a question or completes a statement (Gronlund, 1993).

Triangulation is defined as using either multiple data sources, researchers, theories, or methods to study a single issue (Glesne & Peshkin, 1992; Patton, 1990). In the present study, using interviewing with quantitative survey data is an example of triangulation of data sources.

True/False test items are comprised of a declarative statement that a student must judge as true or false (Gronlund, 1993).

Delimitations and Limitations of the Study

1. Because the data for this study involve one university, and one course within that institution, findings should be interpreted cautiously, within the content-area investigated (i.e., American Heritage or history), and not generalized to other domains (e.g., science, or math).

2. The data collected in the classroom portion of this study are, by necessity, restricted by the American Heritage instructor's current testing practices; therefore, only the testing formats employed in the various exams are used in the statistical analyses, and evaluated by subjects on the S-anxiety measure.

3. Because a history course, such as American Heritage, relies heavily on reading and writing in assessment, speaking proficiency is not being assessed. It should not be overlooked in other content-area courses, such as science, where performance assessments may require students to orally demonstrate their content knowledge.

4. To create a clear and holistic view of learners' opinions, concerns, and preferences in demonstrating content knowledge in a second language, data reduction is inevitable in categorizing and making sense of interview data.
5. Although testing issues permeate all levels of schooling (i.e., primary, secondary, and postsecondary), findings in this university-level study cannot be generalized to each context.

6. Although testing concerns for ESL/LM students are of importance for external and internal examination systems, the findings of this study, with its focus on classroom achievement testing, can only be cautiously generalized to external, standardized, national examinations.

7. The MTELP is used to assess reading, grammar, and vocabulary English language proficiency in this study. It is a standardized language test, and as such does not present a complete record of a ESL/LM student's language ability.

Chapter Summary

The intent of this chapter was to frame the study of ESL/LM students' test performance in content-area courses in an appropriate demographic, educational, and linguistic context. ESL/LM students pose a significant challenge to twenty-first century educators, who must grapple with understanding what constitutes both equal and equitable educational opportunities for diverse student populations. There is consensus among educators that ESL/LM students are at a disadvantage in fully demonstrating their knowledge in content courses. Very little research has focused on delineating the relationship between various test-taker factors—language proficiency, test anxiety, testing preferences, and personal characteristics—and content knowledge in testing situations. The research questions outlined in this chapter are designed to explain second language learners' test performance in content-area classes from both student and classroom perspectives. The following chapter will elaborate and specify further the empirical evidence and theoretical perspectives which inform, motivate, and constrain the present study.
CHAPTER 2

REVIEW OF LITERATURE

This study sets out to explore the relationship between various test-taker variables (i.e., language proficiency, test anxiety, testing preferences, and personal characteristics) and ESL/LM students' test performance in content-area classes. The review of literature which follows considers relevant research and theory from the fields of second language acquisition (SLA), psychology, and classroom testing. The review of literature is divided into several sections. First, the theory and research related to second language (L2) proficiency, test anxiety, and testing preferences (i.e., the purposed independent variables) will be addressed. Second, research related to classroom test performance of ESL/LM students (i.e., the purposed dependent variable) will be reviewed. Third, studies predicting academic success of ESL/LM students will be considered, in that these studies have focused on both language proficiency and performance. Finally, the theoretical perspective interwoven within the review of literature will be concisely summarized. The theory and data chronicled herein provide the rationale for the present study.

Language Proficiency

When the focus is on educating ESL/LM students in primary, secondary, and postsecondary settings, the issue of language proficiency is undeniably at the center of the debate for parents, educators, administrators, policy makers, and politicians. Snow and Brinton (1988, p. 554) acknowledge that, "there exists a threshold level of proficiency without which" ESL/LM students could not succeed in academics. Defining that threshold of proficiency is controversial, and most problematic for students at the primary and
secondary level. Nevertheless, educators at all levels have a vested interest in defining standards and criteria for what constitutes an appropriate level of English language proficiency for participation in academics. Establishing this threshold of proficiency, however, is further complicated by the fact that the field of SLA is in its empirical adolescence, if not infancy.

Despite researchers' interest in L2 proficiency little is definitively known concerning "what is involved in knowing a language," (Shohamy, 1988, p. 165). Lantolf and Frawley (1988) assert that "there is nothing even approaching a reasonable and unified theory of proficiency." (p. 186). There is increasing pressure from inside and outside the SLA field for the development of reliable and valid methods of testing proficiency (Barnwell, 1989). Shohamy (1988) observes, "It is the complexity of the language trait that creates a need for a special discipline called language testing," (p. 165).

Language ability was initially conceived as a unitary trait by Oller in the mid-seventies. He conducted a series of 'g-studies' investigating a "single, global ability," (Bachman, 1988, p. 154). By 1983 this notion was found untenable, and lead to a consensus in the field that language ability is a multidimensional construct (Bachman, 1990; Valdman, 1988; Young, 1992).

The delineation of the components that comprise language abilities has changed substantially over the years. Bachman (1990) notes that early skills/components models put forth by Lado and Carroll in the sixties distinguished between macro- (e.g., listening, speaking, reading, writing) and micro- (e.g., grammar, pronunciation, and vocabulary) language skills, but failed to explain how these various skills were related to each other or influenced by context. In the mid-seventies, Hymes (1972) expanded on Chomsky's distinction between competence and performance to define communicative competence as being comprised of grammatical, sociolinguistic, and contextual competencies. Hymes focused attention on the rules of language use. In the broadest terms, communicative
competence is "the ability to recognize how context constrains communication" (Trenholm & Rose cited in Chen, 1990, p. 9).

Canale and Swain (1980) improved upon this conceptualization by restructuring language competence into four aspects: grammatical, sociolinguistic, discourse (i.e., cohesion in form, coherence in thought), and strategic (i.e., communication and coping strategies). Communicative competence has come to be seen as dynamic, context-specific, and relative, depending on the cooperation of all interlocutors (Savignon, 1983). The interactive nature of these competencies is stressed in the literature (Yalden, 1987b; Savignon, 1983).

This model of communicative competence has roots in various fields, such as psychology, linguistics, communication theory, and pedagogy (Ramirez, 1984; Savignon, 1983). However, Shaw (1992, p. 10) observes that "no one is wholly satisfied with the current model." Several weaknesses have been identified with the notion of communicative competence as a definition of language proficiency.

First, there is difficulty measuring (i.e., the use of global ratings), and establishing independence among competencies (Met & Galloway, 1992; Ramirez, 1984; Shaw, 1992; Young, 1992). Shohamy (1988) points out that the models of communicative competence which have found prominence "mostly list components without examining their relationship," and "have not been validated empirically." (p. 167). Although there is theoretical agreement regarding the multidimensional nature of language abilities, Matthews (1990) cautions against the creation of instruments that "allocate equal marks for the various sub-skills as if the relationship between them was a simple one of addition," (p. 118). Another weakness of the communicative competence model is that it does not account for the possibility of nonnative speakers being more competent than native speakers (Marsh, 1989; Shaw, 1992). Finally, much of what is attributed to
sociolinguistic competence remains unanalyzed and undocumented for those responsible for teaching and testing.

Despite these criticisms, Canale's and Swain's (1980) model of communicative competence is broadly accepted in the SLA field. More recently, Bachman (1988) proposed a new term—communicative language ability (CLA)—for the model, and modified its components slightly. Bachman proposed that three components make up CLA; namely, 1) language competence (grammatical, textual, illocutionary, sociolinguistic), 2) strategic competence (assessing, planning, executing communication goals), and 3) psychophysiological mechanisms of communication (channel, mode). Bachman's (1988; 1990) model was developed specifically to be a guide for language testing, not to be a complete theory of language abilities. Bachman's model has had some impact on how the construct of language ability is investigated and viewed in the context of proficiency testing.

One other significant aspect of language proficiency relates specifically to ESL/LM students in school settings; namely, Cummins' (1980) empirical distinction between what he termed "basic interpersonal communication skills" (BICS) and "cognitive/academic language proficiency" (CALP). For Cummins, CALP replaced Oller's unitary proficiency, and was defined as "those aspects of language proficiency which are closely related to the development of literacy skills in L1 and L2." (p. 177). Academic language abilities, therefore, are considered interdependent in a learner's L1 and L2.

Cummins defined BICS as being aspects of pronunciation, fluency, and sociolinguistic competence, and he considered these skills to be independent of CALP. The components of BICS are not thought to comprise a unitary competence.

The distinction between CALP and BICS is pivotal in the assessment of bilingual and ESL/LM students. Kinsella (1992) reaffirms this point as follows:
more critical to success in secondary and postsecondary schools, however, is academic language... which enables students to deal with cognitively demanding language tasks at school: formal lectures; textbooks in social science, science, and mainstream English classes; and both teacher-constructed and standardized tests, (p. 128).

This difference between BICS and CALP abilities can be puzzling to educators outside the field of SLA. Educators, not familiar with SLA theory and research, may not understand that academic language proficiency takes five to seven years to develop (Collier, 1987). Rosenthal (1996) reminds content-area teachers that, "poor test grades, ungrammatically written assignments and little participation in class discussions do not necessarily mean that ESL students are dumb, lazy, or not studying; rather, these students may still be developing BICS and/or CALP in English," (p. 49).

Cummins (1980) asserts that "natural communication tasks do not assess CALP," (p. 177). At the primary and secondary level, however, Carrasquillo and Rodriguez (1995) report that language proficiency is typically defined solely on the basis of oral language abilities, or home language surveys. As Tikunoff (1985) contends, "oral language proficiency measures provide insufficient data for decision making about the schooling needs of LEP students," (p. 5).

Long (1993) observes that SLA researchers have "an obligation to act as quickly as possible to reach the point where we can respond to practitioners' questions on the nature of language abilities," (p. 229). Although language testing experts debate what constitutes a unified theory of proficiency, the language proficiency of ESL/LM students is being assessed worldwide through various testing alternatives. Which language test is employed depends on the academic context.

Decisions regarding acceptance into various tertiary institutions in the U.S. are most prevalently based on ESL/LM students' performance on norm-referenced, standardized tests. The Test of English as a Foreign Language (TOEFL), the Michigan English Language Assessment Battery (MELAB), or older versions of this test battery
(i.e., MTEL) are commonly used by colleges and universities as measures of English proficiency. These tests are characterized by fixed content, standard administration and scoring procedures, and known reliability and validity properties (Bachman, 1990).

These language tests are used in conjunction with previous grades, letters of recommendations, and other measures of achievement or aptitude to determine an ESL/LM students' readiness for university work (Bachman, 1990). In contrast to the primary and secondary setting, oral proficiency is not a determining factor in a student's acceptance into postsecondary studies.

In summary, determining a student's language proficiency is a critical factor in placing and accepting students in regular content-area courses at the primary, secondary, and postsecondary levels. Pundits have not achieved consensus theoretically or empirically on what level of proficiency is required for academic work, or how to resolve the fact that academic language skills take years to develop. Students cannot afford to wait for academic content until the academic language skills needed develop.

SLA educators have been able to provide admission boards, counselors, and teachers with guidelines for 1) exiting students from ESL programs, and 2) recommending linguistic readiness for university studies based on various English language proficiency exams. SLA researchers have traditionally focused only on language testing issues, not on the impact of language proficiency on content-area testing. Nevertheless, as growing numbers of ESL/LM students compete for postsecondary education opportunities, educators need to understand how language proficiency impacts academic performance; therefore, content-area testing remains a crucial issue for investigation.

**Test Anxiety**

Test anxiety has been investigated empirically as a psychological construct for over eight decades (Spielberger & Vagg, 1995b). As noted in Chapter One, test anxiety is defined as a situation-specific form of general anxiety which centers on a person's fear of failing an
exam (Sapp, 1993). Test anxiety is further defined by several concepts; namely, the
distinctions between trait- and state-anxiety, facilitating and debilitating anxiety, cognitive
and emotional aspects of anxiety, and test wiseness. Following brief discussion of these
concepts, test anxiety and performance will be addressed, especially as it relates to ESL/LM
test takers.

First, in 1958 Cattell and Scheier identified two "distinct anxiety factors": trait- and
state-anxiety (Spielberger, 1966, p. 13). For research purposes, Spielberger argued that
these two types of anxiety "be operationally and conceptually distinguished," (p. 15).
Trait-anxiety was defined as a stable aspect of personality, representing a general
inclination toward anxiety. State-anxiety, on the other hand, was defined as situation-
specific, task-dependent, and transitory. According to Spielberger and Vagg (1995b), the
state-trait view of anxiety is "widely accepted in anxiety research," (p. 8).

Another important contribution to test anxiety research was made by Alpert and Haber
(1960). They developed a measure called the Achievement Anxiety Test (AAT), which
was comprised of two independent scales: 10 items measuring debilitating test anxiety, and
9 items measuring facilitative test anxiety. Prior to Alpert's and Haber's (1960)
delineation of facilitative and debilitative anxiety, test anxiety was believed to have only
detrimental effects on test performance. Their investigation demonstrated that subjects
could possess varying levels of both types of test anxiety. Additionally, the AAT proved to
correlate more strongly with academic performance than other general measures of anxiety.
Facilitating and debilitating anxiety remain well-established psychological constructs. In
terms of trait- and state-anxiety, the AAT is considered a measure of trait-anxiety.

Subsequent research has clarified that S-anxiety in testing also has two major
components: worry and emotionality (Spielberger & Vagg, 1995b). Worry is the
cognitive aspect of state-anxiety, expressed as a person's negative thinking and self-doubts
regarding testing outcomes. Emotionality is the physiological manifestation of state-
anxiety. In particular, research has shown only worry to be strongly associated with poor test performance because worry cognitions divert attention from the test-taking task (Sapp, 1993; Spielberger & Vagg, 1995b).

Finally, Anderson and Sauser (1995) observe that test anxiety and test wiseness are often discussed together, with anxiety explaining poor performance, and test wiseness explaining "better-than-expected performance," (p. 23). In 1965 Millman, Bishop and Ebel defined test wiseness as the "capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score," (cited in Anderson and Sauser, 1995, p. 23). Research has demonstrated that increasing test wiseness can improve test scores, and that high- and low-anxiety students differ significantly in their test wiseness (e.g., various studies cited in Sapp, 1993, or Spielberger & Vagg, 1995b). Being quick and accurate, following directions, guessing, or using clues provided in test items are examples of the types of test-wise skills that improve test performance (Sapp, 1993). Test wiseness is considered a cognitive construct which is "independent of examinees' knowledge of material being tested," (Sapp, 1993, p. 189). In this sense, test anxiety is considered the "emotional component" and test wiseness the "cognitive nature" of test taking (Anderson & Sauser, 1995, p. 23).

The concepts of trait- and state-anxiety, facilitating and debilitating anxiety, worry and emotionality, and test wiseness have each contributed to delineation of the test anxiety construct. Anderson and Sauser (1995), summarizing Hembree's 1988 comprehensive review of test anxiety research, note that test anxiety:

(a) causes poor performance; (b) is inversely related to students' self-esteem; (c) is directly related to students' fears of negative evaluation, defensiveness, and other forms of anxiety; (d) is influenced by ability, gender, and school grade level; and (e) can be reduced effectively by a variety of treatments. (p. 15).

Madsen and Murray (1984) also note that students who are prone to anxiety in testing situations have a "unique profile in terms of their reactions to exams and exam conditions."
(p. 16). For example, females are more prone to anxiety in testing than males, and low anxiety subjects "score better on many different measures" than high anxiety subjects (Madsen & Murray, 1984, p. 1). Furthermore, test anxiety has been shown to affect minority students to a greater degree than majority students (e.g., Phillips, Pitcher, Worsham, & Miller, 1980; Sapp, 1993). Differences in the cultural background among minority students' can also contribute to how students affectively react to tests (Scott, 1986).

Many studies describe the connections between sociolinguistic background, test anxiety—in particular state-anxiety as a good predictor of performance (Becker, 1983)—and test performance. For example, Phillips et al. (1980) found that for Mexican-American females anxiety was the best predictor of reading and math achievement. Diaz-Guerrero (1976), who studied Mexican and American children, observed that Mexican children struggle on exams because "most of their responses are designed to please adults... On the other hand, the North American culture, with its stress on competitive achievement and self-sufficient individualism, is more consonant with objective exams," (p. 141). Diaz-Guerrero (1976) also found socioeconomic factors to be important predictors of test anxiety: "subjects from lower socioeconomic levels consistently score higher in anxiety," (p. 141).

Madsen and Murray (1984) applied facilitating and debilitating anxiety constructs to content-area testing and language testing. They carried out two pilot studies with students at Brigham Young University (BYU): one with nine matriculated, graduate students (native and non-native English speakers) taking TESOL language testing and methodology courses (i.e., Study One); the other with seventeen ESL students in an intensive English program setting (i.e., Study Two).

In both studies, Madsen and Murray (1984) were interested in students' trait- and state-anxiety. They used Alpert's and Haber's 19-item AAT to measure facilitating and
debilitating trait-anxiety. To measure state-anxiety, Madsen and Murray designed a one-page questionnaire inquiring about the difficulty of each part of the exams (e.g., multiple choice, short answer, essay).

In Study One, Madsen and Murray found that native-English speakers showed less test anxiety than ESL graduate students, and did not show preference for any particular test format; however, the ESL students "felt they could perform more successfully on objective items than on completion and essay questions," (1984, p. 6). Madsen and Murray also found that high trait-anxiety students showed concern over the length of a test, and strong preferences for 1) loose time constraints in testing; 2) previewing a test before beginning; and 3) group study over individual study.

In Study Two, which focused on students taking BYU's English Language Center (ELC) Placement Battery, Madsen and Murray (1984) found, contrary to expectations, that high anxiety subjects outperformed those with low anxiety "on every section of the battery," (p. 13). While high and low anxiety subjects agreed that the reading and writing portions of the language exam were the most difficult, they disagreed in their assessment of the listening, speaking, and grammar portions of the exams. High anxiety subjects considered these sections difficult; low anxiety subjects considered them easy.

In an earlier study, Madsen (1981) specifically explored gender, language background, and language proficiency factors related to state-anxiety among ESL students (N = 114) taking the ELC Placement Battery. Gender and language background variables proved statistiically significant, with females and Japanese students indicating more state-anxiety in testing than males and Spanish speakers.

Overall, these studies suggest that test anxiety, in particular trait- and state-anxiety measures, should be taken into account in interpreting and comparing test performance among ESL/LM students in various exam situations. Test anxiety, as a factor, appears to have potential in illuminating ESL/LM students' test performance in content-area classes.
Testing Preferences

Every person approaches a testing situation with their own assumptions, concerns, and expectations (Sarason, 1980). For many ESL/LM students, their native linguistic, cultural, and educational knowledge about testing does not match the testing practices they face in their second languages. This disparity often means that test performance does not accurately represent an ESL/LM student's actual content knowledge.

As noted in Chapter One, psychometricians have traditionally paid little attention to test takers or their preferences in classroom testing (Zeidner, 1987). Nevertheless, several recent studies with native English speakers suggest that test takers can provide useful insights into testing processes. Following discussion of these native speaker studies, research into ESL/LM students' testing preferences will be reviewed. For clarity across these studies, Gronlund's (1993) definitions of true/false, short answer/completion, multiple choice, restricted response, and extended essays will be used, as presented in Chapter One.

Nield and Wintre (1986) found language-majority university students most preferred taking tests with restricted-response questions, followed in order of preference by essay, multiple-choice, fill-in-the-blank, and true/false questions. Zeidner's (1987) study with junior high school students revealed that "essay exams were perceived to be more fair than multiple-choice exams," because essay exams allowed students "the opportunity of accurately and optimally expressing their knowledge and ideas in writing," (p. 355). These student perceptions of fairness and validity of essay questions are also confirmed in research by Gellman and Berkowitz (1993), and Nield and Wintre (1986).

Although language-majority students feel essay exams allow them to demonstrate their knowledge more accurately, studies also show that students prefer objective test items (e.g., multiple choice) over subjective items (e.g., essay) for very specific reasons. For example, junior high school students believed that they had "a better chance of succeeding
on multiple-choice... exams." (Zeidner, 1987, p. 355). University students found objective items advantageous because "answers were provided (which allowed them to guess) . . . and required little mental or physical effort," (Nield & Wintre, 1986, p. 197). Gellman and Berkowitz (1993) found that students preferred objective items because they could "get a good score with limited knowledge," (p. 24). Furthermore, they suggested that language-majority students "do not necessarily prefer the measure that is perceived as most likely to demonstrate the extent of their knowledge. . . rather, they prefer the type of test on which they believe it is easier to do well," (p. 18).

Foos (1992) carried out a study with language-majority university students (N = 84) investigating the interaction between expected test form (e.g., multiple choice vs. essay) and expected difficulty (e.g., easy vs. difficult). Foos found that when students were told to expect a difficult and/or essay test, they performed better than when they expect easy and/or a multiple-choice test. In addition, results suggested that students believe that "any test, even an essay test, could be easy but that a multiple-choice test can never be very difficult," (p. 209).

Research with language-majority students also seems to indicate that gender is an important variable in understanding test-format preferences. Gellman and Berkowitz (1993) found that women strongly preferred essay items over multiple-choice items. Anderson (1989) asserts, "women are less likely to gamble when they are really unsure of the answer," (p. 175). A study by Bridgeman and Lewis (1994), using Advanced Placement Examinations (AP), found that while essay scores were nearly equivalent for females and males, males significantly out scored females on multiple-choice sections. In particular, Bridgeman and Lewis asserted that "a greater reliance on multiple-choice scores" would disadvantage females in receiving college credit or advanced placement (p. 48).

Although Kinsella (1992) observes that ESL/LM students need strategies for coping with both objective and subjective test questions, few studies have investigated the testing
preferences of ESL/LM students for demonstrating their content knowledge. Some recent articles, however, confirm a growing interest in content-area testing from an ESL/LM perspective (e.g., Harklau, 1994; Hayward, 1990; Horowitz, 1986; Kinsella, 1992; Leki and Carson, 1994; Leki, 1995). A review of these articles follows.

Leki and Carson (1994) note that writing skills are important for "achieving academic success . . . because many courses evaluate students through some form of written text (e.g., essay exams, short-answer essays, research papers)," (p. 82-83). Because ESL/LM students tend to be "insecure when expressing themselves in writing," (Madsen & Murray, 1984, p. 7), certain test formats, such as essay or restricted-response questions, pose special problems for ESL/LM students (Horowitz, 1986; Kinsella, 1992; Leki, 1995).

Horowitz points out that production-type test items require students to "find, organize, and present data according to fairly explicit instruction," (p. 455). Hayward (1990, p. 754) underscores the importance of both language ("vocabulary, syntax, organization, and rhetoric) and content ("a grasp of the material under discussion") in answering essay questions. In a study by Leki and Carson (1994), a subject noted these same concerns for precise vocabulary and speed as follows:

In a short essay, I had to use the words that should have very specific meaning. Since my word choice is limited, I have almost always hard time to find those words; therefore, my essays cannot make good arguments but rather be very simple, (p. 94).

Kinsella (1992) also offers that ESL/LM students may fail to interpret test directions correctly for a lack of academic language ability. For example, an essay prompt asking students to "Trace the early waves of immigrants to the U.S." my lead an ESL/LM student to respond to "early immigration," rather than the "key direction word trace ," (p. 127). Furthermore, Chiste and O'Shea (1988) found that ESL/LM students need to be warned against selecting essay questions "by length alone," and need to be encouraged to "read all questions before making a selection," (p. 683-684).
Objective test items may also pose difficulties for ESL/LM students. Leki and Carson (1994, p. 94) stated that ESL students require a "disproportionate amount of time" for just reading and selecting correct answers on multiple-choice items. Harklau (1994), on the other hand, found ESL/LM students "proficient in bluffing their way through such mechanical exercises without a clear idea of what they were talking about," (p. 254).

In a qualitative study of five matriculated, international ESL students, Leki (1995) found that each student dealt with the essay writing task on exams differently. One Taiwanese subject approached her professors prior to exams for clarification on how to prepare for the essay exam; a French subject relied on a L1 formula--thesis-antithesis-synthesis--to succeed. Whatever strategies ESL/LM students employ in responding to essay questions, it is clear that these strategies "are only part of the vital repertoire" of skills ESL/LM need to succeed on content-area tests.

In summary, research reveals clear testing preferences for language-majority students. From a test-taker perspective, perceived ease and difficulty, perceived potential for success, and gender are all variables influencing test preference and performance. Descriptive data from ESL/LM students intimate that testing preferences and performance are closely and differentially linked to language proficiency, specifically to vocabulary, writing, and reading skills. These inter-relationships, however, are not empirically substantiated. More qualitative and quantitative probing of these issues are justified given the increasing concern over ESL/LM student success in mainstream, content-area courses.

**ESL/LM Test Performance**

Circumventing the questions raised by the use of various national testing programs with ESL/LM students is a difficult task for administrators, policy makers, and teachers (LaCelle-Peterson & Rivera, 1994; Lam, 1993; Madaus, 1994). Lam (1993) observes that developers of standardized test usually make five assumptions about test takers; namely, that test takers have 1) no linguistic barriers to hamper performance; 2) appropriate content
knowledge; 3) experience or sophistication taking tests; 4) motivation to do well; and 5) no significant test anxiety. Many, if not all, of these assumptions are violated when ESL/LM students are subjected to standardized achievement tests, and certainly the reliability and validity of such tests are reduced.

Classroom teachers, faced with the same issues on a local level, are implementing modifications in their testing practices with ESL/LM students. Brinton et al., (1989) list timed tests, long readings, poor or tricky distractors, essays, research papers, and oral presentations as "evaluation tasks which often disadvantage" ESL/LM students (p. 184).

Latulippe and Light (1983) describe the skills students must have for succeeding on content-area tests as follows:

In the university, our students need to do more than write unified, coherent paragraphs. They must be able to write rapidly, paraphrase ideas that they hear in lectures as notes, and use these notes to write clear, essay question answers within the time limit of a fifty minute class, (p. 8).

Bernhardt et al., (1995), Brinton et al., (1989), Kinsella (1992), LaCelle-Peterson and Rivera (1994), and Rosenthal (1996) each list techniques for improving ESL/LM students' test performance. Chief among these suggestions are the following: doing away with time restraints, allowing bilingual dictionary use, using multiple indicators, using more than writing skills in assessment, separating language and content in grading practices, using pictures and diagrams, being flexible, and teaching test-taking skills.

Although the aim of ESL programs, specifically EAP programs, is to prepare ESL/LM students for the linguistic and academic challenges of content-area courses (Chamot & O'Malley, 1987; Ferris & Tagg, 1996; Spack, 1988), complete preparation is difficult to accomplish (Barker & Bohlman, 1991; Carrasquillo & Rodriguez, 1995). Referring to the years it requires to develop academic language abilities, Rosenthal (1996) notes, "five to seven years is a considerably longer period of time than that either allotted to any ESL program or to the earning of most college degrees," (p. 19).
Research has shown that ESL/LM students with test wiseness perform better those who lack familiarity with U.S. testing practices (Dunkel, 1988; Gellman & Berkowitz, 1993; Lam, 1993; Madsen & Murray, 1984; Soto, 1992). Difficulties in testing, however, are common concerns among ESL/LM students (Hirsch, 1988; Light and Teh-Yuan, 1991; Smoke, 1988). In surveying ESL students regarding how well their EAP writing courses prepared them for university work, Leki and Carson (1994) found that students expressed a need to improve their vocabulary knowledge, increase their speed in language processing, and work on other grammatical aspects of writing. Overall, Leki and Carson summarized that students were frustrated:

(a) with the amount of time it takes them to access appropriate lexical and grammatical forms, (b) with the imprecision with which they were forced to express themselves. . . . and (c) with the gap between the writing. . . in EAP classes and the writing they are then required to do in other courses, (p. 95).

Short (1993) and Madaus (1994) both mention the potential of performance, alternative, or authentic assessment procedures to provide educators with more accurate information about ESL/LM students' language and content knowledge. A study by Bernhardt et al. (1995) investigated performance assessments for 20 fourth and fifth graders in a Spanish immersion science classroom. Students were asked to read three expository and one narrative text in Spanish, their second language. Using reading recall procedures, students recalled information from each text using the language of their choice: English or Spanish. In addition, an appropriate performance assessment was carried out.

Bernhardt et al., (1995) found significant correlations between literacy measures (i.e., reading recall and the California Achievement Test); they found no relationship between reading scores and performance assessment outcomes. Bernhardt et al. (1995) concluded that, "performance assessments clearly tap knowledge other than that which is tapped in reading recall and standardized achievement test measures. What that other knowledge is, is certainly not discernible from our data, but needs to be investigated," (p. 6). To date, there is little empirical evidence to suggest that alternative assessments are any more
equitable, or any less "corruptible" than traditional standardized tests for ESL/LM students (Madaus, 1994, p. 84). Short (1993) laments that "no approach is without drawbacks," (p. 653).

Finally, regarding ESL/LM students' test performance, research has demonstrated that content-area faculty members consider both language and content in grading essay responses (Santos, 1988; Vann, Meyer, & Lorenz, 1984). Santos and Vann et al. found that faculty in the sciences were less tolerant of, and more irritated by language errors than faculty in other disciplines. Although Santos (1988) noted that professors were willing to "judge content and language independently," they were not able to overlook lexical errors which were found to "impinge directly on content," (p. 84). According to Horowitz (1986), the present focus on process writing, with multiple drafts, has left ESL/LM students "unprepared for essay examinations," (p. 446).

In summary, the test performance of ESL/LM students is influenced by language proficiency, test anxiety, testing preferences, and other test-taker characteristics. These factors can jeopardize performance on national, state, or classroom exams. The challenges associated with making testing equitable for ESL/LM students in education are enormous. Some educators are altering their testing practices to accommodate ESL/LM students' needs; however, little is empirically known about second language learners' testing preferences and performance in content-area courses. The need for baseline data is intimated. As Hernández (1992, p. 144) suggests "the integration of insights provided by both quantitative and qualitative methodologies," is needed to understanding ESL/LM students' test performance.

**Predicting Academic Success**

There is a substantial body of research aimed at predicting ESL/LM students' academic success, especially at the primary and tertiary levels. This research clearly focuses on the
relationship between language proficiency and academic success. Academic success, however, is defined differently for each study.

At the primary level, proponents of bilingual education and English-Only legislation are polarized on the role of students' native language (L1) in U.S. schooling. Faltis and Hudelson (1994) admit that prior to 1990 research and information on bilingual education in primary and secondary settings were scarce. However, present advocates of bilingual education, such as Collier (1992), are armed with empirical evidence to support their position on the efficacy of a bilingual approach in promoting academic success among ESL/LM students. In summarizing the results of nineteen studies, Collier maintains the following:

the greater the amount of L1 instructional support for language-minority students, combined with balanced L2 support, the higher they are able to achieve academically in L2 in each succeeding academic year, in comparison to matched groups being schooled monolingually in L2, (p. 205).

Findings such as these have not silenced skeptics of bilingual education or advocates of English-Only legislation. On a national front, bilingual education is "under attack," and threatened by "deep cuts" in federal support (Schnailberg, 1996a, p. 10). Schnailberg (1996b) even documents unrest among Hispanic parents in New York, whose children participate in bilingual education programs. These parents say they "don't want to be part of an 'experiment' where maybe their kids will come out having learned English in six years," (1996b, p. 11). At the primary level, the role of English in education remains controversial.

At the tertiary level, English language proficiency is a prerequisite for admission into a college or university. Graham's (1987) review of 18 studies which predicted academic success based on L2 language proficiency revealed inconclusive results, as discussed in Chapter One. These prediction studies calculated correlations between English language proficiency, as measured by a standardized language exam such as the TOEFL (e.g.,
Johnson, 1988), and academic success. In a majority of studies, academic success was defined by high school or college grade point average (GPA); however, verbal or quantitative scores on the Graduate Record Exam (GRE), obtaining a degree, or class grades have also been utilized (Graham, 1987). For 14 of the studies Graham reviewed, half revealed strong correlations between language proficiency and academic success, and half did not. The four remaining studies were inconclusive.

The contradictory findings of these prediction studies suggest several weaknesses in design. First, GPA itself is an inexact measure, and difficult to interpret with accuracy. For example, GPA is influenced by differences among instructors, course content, methods of grade calculation, and institution factors. A student's GPA in one institution may not have the same meaning in another. Graham (1987) cautions that GPA "is not always a valid indicator of academic achievement," (p. 506). Light, Xu, and Mossop (1987, p. 259) suggest that it may not even be "the most important criterion" for success.

A second weakness lies with the use of commercial measures of English language proficiency. Graham (1987, p. 507) notes that proficiency in prediction studies is defined as "performance on the test." This view of proficiency does not match the current focus on communicative competence, with its interest in language use in context. Graham asserts that the use of multiple-choice language exams as measures of proficiency is questionable; however, she admits "it has yet to be conclusively demonstrated that traditional tests do not predict language performance," (Graham, 1987, p. 507).

Performance scores on language proficiency exams are usually considered holistically in prediction studies; however, there is some evidence that "vocabulary knowledge is the single most important" factor "in relation to academic achievement," (Leki and Carson, 1994, p. 95). Graham (1987) reports the same finding for a study by Saville-Troike. Studies by Delisle (1982), Khalil (1985) and Sheorey (1986) each confirm the important role of vocabulary, over grammar, in oral and written communication.
According to Vandrick (1995), Hernández (1992), and Graham (1987), a third concern with prediction studies is properly defining the ESL/LM population. Geisinger and Carlson (1992) clarify that "while we may view Mexicans, Puerto Ricans, Cubans, and Central Americans all as Hispanics, each group has significant demographic, behavioral, and geographic differences," (p. 1).

Rosenthal (1996) and Vandrick (1995) believe that international students, who often come from privileged backgrounds, and immigrant/refugee students should be recognized as different populations. While some language minority groups are "marginalized, and disempowered," others will return to their countries "and step into positions of power, wealth, and influence." (Vandrick, 1995, p. 375).

Although little research has focused on distinguishing between groups within ESL/LM populations, Pearson (1993) carried out a prediction study targeting only English-dominant "academically well-qualified Hispanic bilingual students," (p. 344). Scholastic Aptitude Test (SAT) scores were used to predict college success for Hispanic and White students who were equivalent on socioeconomic and cultural variables. Although Hispanic students scored approximately 45-points below their White peers on the SAT, these students were found to be "academically equivalent . . . except in terms of SAT scores," (p. 353). Pearson considered these results peculiar to the specific group of bilingual students studied, and "not relevant for the evaluation of most international students," (p. 352).

Finally, Graham (1987) concludes her review of prediction studies by advocating the investigation of other factors, along with language proficiency, to predict academic success. Among the other possible factors considered important to academic success are personality, attitudes, motivation, cultural adaptation, family background, socio-economic status, class standing, and self-ratings (Graham, 1987; Johnson, 1988; Light & Teh-Yuan, 1991; Light et al., 1987; Wright & Michael, 1989).
In summary, prediction studies have failed to conclusively correlate language proficiency with ultimate academic achievement. While many studies have focused on predicting success using polluted GPA scores, few have attempted to explain academic success in a classroom setting, using English language proficiency and teacher-specific test grades. Research designs which 1) consider language proficiency holistically and as separate skills (i.e., reading, writing, grammar, and vocabulary); and 2) relate proficiency along with other factors, such as test anxiety and testing preferences, to test performance in a particular class would make a valuable contribution to understanding the academic success of ESL/LM students.

Theoretical Perspective

In investigating various ESL/LM test-taker variables and performance on content-area tests, theories from the fields of SLA, psychology, and educational testing provide the theoretical foundation for this study. Although the theoretical contribution of each field has been presented throughout the review of literature, for clarity these contributions, as related to the present study, will be succinctly reiterated.

SLA researchers have demonstrated that language proficiency involves a concern for grammar, sociolinguistic appropriacy, discourse, and various strategies (Bachman, 1990; Canale & Swain, 1980; Hymes, 1972; Savignon, 1983). Specifically, the distinction between general and academic language proficiencies has also been significant in understanding ESL/LM students' academic performance in content-area classes (Collier, 1987; Cummins, 1980). As applied to the present study, this SLA conceptualization of language proficiency holds that ESL/LM students' language proficiency influences test performance, because students are forced to demonstrate content knowledge in their second language, which is necessarily more restricted or limited than their first language proficiency.
Test anxiety is a well-researched construct in psychology (Sapp, 1993; Spielberger, 1966; Spielberger & Vagg, 1995). Although test anxiety has been investigated most prevalently among language-majority students, research does demonstrate that test anxiety is also a language-minority issue (Phillips et al., 1980; Sapp, 1993). Test anxiety theory indicates that trait- and state-anxiety impact test performance. In the present study, this theory predicts that measures of ESL/LM students' test anxiety will influence test performance because language-minority students may be less test wise than their language-majority peers (e.g., Diaz-Guerrero, 1976).

Classroom testing experts acknowledge that test format is an important consideration in test development (Gronlund, 1993; Zeidner, 1987). There is also growing interest in test-takers' characteristics and format preferences in regard to test development (Gellman & Berkowitz, 1993; Nield & Wintre, 1986; Zeidner, 1987). Similar interests are developing among educators working with ESL/LM students (Bachman, 1990; LaCelle-Peterson & Rivera, 1994; Lam, 1993; Madaus, 1994). Classroom testing research suggests that learners have preferences for demonstrating their content knowledge in testing situations. These preferences influence test performance when students are not allowed to demonstrate their content knowledge according to their testing preferences, or in the format that allows them to get the highest score.

These theories suggest that language proficiency, test anxiety, and testing preferences are potentially important independent variables in accounting for ESL/LM students' test performance in content-area classes. The literature also suggests that various test-taker characteristics, such as gender and language background, may be important variables in relationship to anxiety and preferences in testing. SLA, psychology, and testing theory each purport that there is value in approaching test performance in content-area classes from an individual perspective, taking into account individual learner's proficiency, anxiety, preferences, and personal characteristics. While causal relationships are not explored, there
is evidence to suggest that these variables have potential in accounting for ESL/LM students' test performance.

The rationale for designing and conducting this study is represented in the following statement: If information about an ESL/LM student's a) language proficiency, b) level of test anxiety, c) testing preferences, and d) personal characteristics are known, then the association of these variables to test performance in content-area classes can be understood my fully. Figure 2.1 represents the variables and interrelationships being investigated in the present study. The data collection and analyses procedures presented in Chapter Three will allow investigation of the relationships among these variables.

**Overall Conclusions**

Given the theoretical, empirical, and practitioner issues raised in the preceding review of literature, several conclusions are warranted concerning content-area test performance from a second language perspective. First, exploration of the language and content dilemma in testing is a pressing concern for ESL and content educators alike. A student's English language proficiency is central in educational decisions regarding participation in content-area courses and testing. Because academic language proficiency takes as many as seven years to develop in a second language, students cannot afford to wait for content if they hope to remain competitive with their language-majority peers. Remaining competitive, however, also means ESL/LM students are jeopardized, to some degree, when demonstrating their content knowledge in their second language.

Second, the increasing number of ESL/LM students in content-area courses is placing new demands on faculty, who are unaccustomed to dealing with language issues in teaching and testing. Collaboration in research and teaching among ESL and content-area faculty appears inevitable and increasingly beneficial.

Third, substantial research effort has been expended to predict the academic success of ESL/LM students in university studies. Because predictions studies using GPA have
Figure 2.1: A visual model of the variables which are potentially important in predicting ESL/LM students' test performance in content-area classes.
yielded ambiguous findings, attempting to predict academic success in a particular course appears to be a viable research alternative.

Fourth, the national, local, and classroom concerns of educators who test ESL/LM students in content-areas are substantial. Current testing practices yield results of questionable reliability and validity for ESL/LM students. Extensive research is needed to clarify what characteristics of tests and test-takers impact content-area testing outcomes for ESL/LM students.

Fifth, theory suggests that language proficiency, test anxiety, and testing preference are potentially significant variables in accounting for variance in the test performance of ESL/LM students in content-area classes. As suggested by the studies reviewed, personal characteristics of test takers may also contribute to explaining ESL/LM students' test performance (i.e., gender, language background, length of residency, length of U.S. university experience, familiarity with U.S. examination system, number of previous American history courses taken).

Sixth, the use both qualitative and quantitative research methodologies is merited when investigating ESL/LM students' test performance in content-area classes.

The theory, research, and conclusions presented in this chapter inform and justify the present study of ESL/LM students' test performance in content-area classes from a second language perspective. Chapter Three provides a description of the research design and methodology employed in carrying out the present study.
CHAPTER 3

RESEARCH DESIGN AND METHODOLOGY

Introduction

In order to investigate content-area testing from a second language perspective, both quantitative and qualitative data were gathered. These data were analyzed to determine what, if any, associations exist among the endogenous variables of language proficiency, test anxiety, testing preferences, and test performance for ESL/LM students in content-area courses. In addition, the influence of gender and language background (i.e., exogenous variables) on test anxiety and testing preferences were investigated.

The present study theoretically adheres to quantitative methods, with the inclusion of a complementary qualitative component. The four stages of data collection represent simultaneous triangulation of data sources. Stages One and Four of this study focused on obtaining an insider's perspective--a first hand account--of the opinions, concerns, and preferences of ESL/LM students in testing by conducting semi-structured interviews. These qualitative data provide linguistic description of ESL/LM students' experiences in demonstrating content knowledge in their second language. These data were gathered to add breadth to the quantitative portion of this study. Stages Two and Three, which rely mainly on survey methods, provide the numeric or quantitative description of this same testing phenomenon.

The purpose of this chapter is to define, specify, and provide the rationale underlying various aspects of the research design and methods used in this study. Following a general discussion of non-experimental research and qualitative research, the subjects, variables,
instruments, data collection procedures, and statistical analyses applied to this study are presented. In addition, the pilot testing of instruments is discussed.

**Nonexperimental Research**

Nonexperimental designs, in contrast to experimental and quasi-experimental designs, lack both treatment manipulation and randomization. Nonexperimental designs are also described in terms of methods as survey research, and in terms of analyses as correlational research (Pedhazur & Schmelkin, 1991). Pedhazur and Schmelkin note that the aim of nonexperimental designs are either predictive (i.e., criterion-related validation) or explanatory (i.e., hypothesis testing). Hatch and Lazaraton (1991) observe that this design is the most common type used among applied linguists, who often are searching for "what is going on" rather than "what caused this." (p. 100).

The present study, which is based on a nonexperimental design, sets out to predict a sample of ESL/LM students' content-area test performance (i.e., the criterion or dependent variable) based on information from other variables (i.e., predictors or independent, such as language proficiency, test anxiety, and testing preferences). In addition, differential prediction is of interest when focus includes such mediating variables as gender or language background. As Pedhazur and Schmelkin (1991) suggest, criterion-related validation is a process concerned with "the degree of successful prediction of a criterion, regardless of whether or not it is possible to explain the process or processes leading to the phenomenon that is being predicted," (p. 32).

Much of the data gathered in this study was obtained using a questionnaire. Survey designs allow researchers to obtain numeric descriptions of a sample of people by asking questions (Creswell, 1994; Fowler, 1993). In this type of design the sample frame, question design, and interview methodology are of central importance.

Given the research questions posed in this study, a survey approach to data collection was preferred for several reasons. First, Brigham Young University (BYU) functions on a
semester system, and a student's test performance in a content-area course (i.e., the dependent variable) is defined by this 16-week period. A survey design allowed for the necessary data to be collected quickly within this time frame. Second, survey methods allowed the maximum amount of relevant data to be parsimoniously collected from one sample group of ESL/LM students. Third, survey methods allowed identification of the important characteristics (i.e., testing opinions, concerns, and preferences) of ESL/LM students being tested at the university setting. Fourth, the classroom-based, group, and face-to-face administration of the various questionnaires assured high response rates, low costs, and low attrition rates during the semester of study. While the language proficiency, T-anxiety, and testing preference data were collected at one point in time, the S-anxiety and test performance data were collected over the course of a semester.

Overall, the quantitative data collected by questionnaires (e.g., testing preferences and T-anxiety), and by particular test performance (e.g., language proficiency scores, and content-area test scores) allowed for the research questions to be answered in a convenient, timely, standardized, reliable, and cost-efficient manner.

Qualitative Research

Qualitative research relies on non-numerical data to describe and understand human experience (Creswell, 1994; Glesne & Peshkin, 1992). Qualitative traditions have emerged from such fields as anthropology, philosophy, sociology, and psychology (Patton, 1990). Qualitative inquiry relies on assumptions which distinguish it from quantitative inquiry (Bogdan & Biklen, 1992; Creswell, 1994; Guba & Lincoln, 1989; Marshall & Rossman, 1995). Among these assumptions is a focus among qualitative researchers on descriptions, processes (not products), natural settings, the emic construction of personal meaning, data collection via human researchers, and inductive theory building. Because of the personal—possibly intrusive—nature of interviewing, ethical issues, such as reciprocity, confidentiality, and informed consent, guide the qualitative researcher (Patton, 1990).
In the present study, the use of semi-structured interviewing techniques, with open-ended questions, allowed exploration of testing, specifically content and language issues in testing for ESL/LM students. Because an investigation of content-area testing from a second language perspective is still exploratory, and current testing theory may not apply to ESL/LM student populations, it is beneficial to consider the test-takers themselves in the testing process. For these reasons, an interviewing component in data collection was considered useful in adding perspective and a gestalt to content-area testing for ESL/LM university students.

Participants

Qualitative Interview Subjects.

The 13 subjects selected for participation in the semi-structured interviews (Stage One) were ESL/LM students enrolled in four-year universities: BYU and The Ohio State University (OSU). Subjects were selected on the purposeful criterion of typicality, and interviewing ceased when no new information was obtained; redundancy was the primary criterion for ending subject selection (Lincoln & Guba cited in Patton, 1990).

Table 3.1 presents the characteristics of the general interview subjects. As shown, subjects represented both genders (i.e., females=6; males=7), and different language backgrounds (i.e., Belarusian, Russian, Portuguese, Spanish, Korean, Chinese, and Arabic). Only two of the thirteen subjects were graduate students. Subjects who participated in the interviews have lived in the United States between seven months and seventeen years, and have been university students between one month and two-and-a-half years. Their university majors included business, linguistics, broadcast communications, architecture, chemistry, computer engineering, musicology; two subjects had not declared majors. These subjects' were selected to elucidate ESL/LM students' opinions, concerns, and preferences in content-area test situations (i.e., Research Question 1).
<table>
<thead>
<tr>
<th>ID</th>
<th>Institute</th>
<th>Major</th>
<th>Class Standing</th>
<th>Gender</th>
<th>Native Language</th>
<th>Country</th>
<th>Years in the USA:</th>
<th>Univ:</th>
</tr>
</thead>
<tbody>
<tr>
<td>01</td>
<td>BYU</td>
<td>Business</td>
<td>U*</td>
<td>M**</td>
<td>Belarusian</td>
<td>Belarus</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>02</td>
<td>BYU</td>
<td>Linguistics</td>
<td>G</td>
<td>F</td>
<td>Portugese</td>
<td>Brazil</td>
<td>1.00</td>
<td>0.08</td>
</tr>
<tr>
<td>03</td>
<td>BYU</td>
<td>Linguistics</td>
<td>U</td>
<td>M</td>
<td>Korean</td>
<td>Korea</td>
<td>3.00</td>
<td>2.50</td>
</tr>
<tr>
<td>04</td>
<td>BYU</td>
<td>Broadcasting</td>
<td>U</td>
<td>M</td>
<td>Spanish</td>
<td>Mexico</td>
<td>0.67</td>
<td>0.33</td>
</tr>
<tr>
<td>05</td>
<td>BYU</td>
<td>Musicology</td>
<td>G</td>
<td>F</td>
<td>Russian</td>
<td>Israel</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>06</td>
<td>OSU</td>
<td>Business</td>
<td>U</td>
<td>M</td>
<td>Russian</td>
<td>Russia</td>
<td>4.00</td>
<td>2.00</td>
</tr>
<tr>
<td>07</td>
<td>OSU</td>
<td>International Business</td>
<td>U</td>
<td>F</td>
<td>Arabic</td>
<td>Jordon</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>08</td>
<td>OSU</td>
<td>Architecture</td>
<td>U</td>
<td>M</td>
<td>Spanish</td>
<td>Puerto Rico</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>09</td>
<td>OSU</td>
<td>Computer Engineering</td>
<td>U</td>
<td>F</td>
<td>Chinese</td>
<td>Taiwan</td>
<td>3.00</td>
<td>0.08</td>
</tr>
<tr>
<td>10</td>
<td>OSU</td>
<td>Undecided</td>
<td>U</td>
<td>F</td>
<td>Spanish &amp; English</td>
<td>Mexico &amp; USA</td>
<td>17.00</td>
<td>0.17</td>
</tr>
<tr>
<td>11</td>
<td>OSU</td>
<td>Business</td>
<td>U</td>
<td>M</td>
<td>Spanish</td>
<td>Puerto Rico</td>
<td>1.00</td>
<td>0.25</td>
</tr>
<tr>
<td>12</td>
<td>OSU</td>
<td>Undecided</td>
<td>U</td>
<td>F</td>
<td>Korean</td>
<td>Korea</td>
<td>4.00</td>
<td>0.17</td>
</tr>
<tr>
<td>13</td>
<td>OSU</td>
<td>Chemistry</td>
<td>U</td>
<td>M</td>
<td>Chinese</td>
<td>Hong Kong</td>
<td>6.00</td>
<td>0.75</td>
</tr>
</tbody>
</table>

*U = undergraduate; G = Graduate. ** M = Male; F = Female.

Table 3.1: Description of General Interview Subjects

47
Course Subjects.

Because the described aim of this study is to explore content-area testing from a classroom perspective, subjects for Stages Two and Three were purposefully selected based on their enrollment in the two international sections of an American Heritage course offered through the History Department at BYU during Fall semester, 1996. BYU is a private, religious, four-year university with an ESL/LM population of approximately 1,800 matriculated students.

Although 62 students enrolled in these American Heritage sections, only 38 students were ESL/LM students. There were 20 males, 18 females, and 15 different languages represented among these students. Based on these languages, students were divided into romance (n = 21), and non-romance (n = 17) language groups. The romance language group was comprised of Spanish (14), French (5), Portuguese (1), and Italian (1) speakers. The non-romance language group contained Arabic (3), Chinese (3), Japanese (2), Korean (1), Cambodian (1), Russian (1), Belorussian (1), Polish (1), Czech (1), Dutch (2), and Norwegian (1) speakers. Only six of these subjects were permanent residents. These ESL/LM subjects had an average length of residency of 3.56 years (SD = 3.15), and an average university experience of 1.64 years (SD = 1.34). The average number of American history courses students had taken previously was 0.42 (SD = 0.60).

These 38 subjects provided data to answer the research questions regarding the role of language proficiency, test anxiety, testing preferences, and personal characteristics (e.g., gender and language group) in predicting test performance in these American Heritage classes. Finally, ten percent of these ESL/LM subjects (n = 4: 3 females; 1 male) were randomly selected for an end-of-term interview regarding their testing experience in American Heritage 100. Table 3.2 presents a description of the four American Heritage subjects interviewed at the end-of-the-term.

48
<table>
<thead>
<tr>
<th>ID</th>
<th>Institute</th>
<th>Major</th>
<th>Class Standing</th>
<th>Gender</th>
<th>Native Language</th>
<th>Country</th>
<th>Years in the USA:</th>
<th>Univ:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BYU</td>
<td>Business</td>
<td>U*</td>
<td>F**</td>
<td>Spanish</td>
<td>Ecuador</td>
<td>2.33</td>
<td>1.50</td>
</tr>
<tr>
<td>2</td>
<td>BYU</td>
<td>Computer Science</td>
<td>U</td>
<td>M</td>
<td>Spanish</td>
<td>Mexico</td>
<td>1.50</td>
<td>1.50</td>
</tr>
<tr>
<td>3</td>
<td>BYU</td>
<td>International Relations</td>
<td>U</td>
<td>F</td>
<td>Italian</td>
<td>Italy</td>
<td>1.50</td>
<td>0.42</td>
</tr>
<tr>
<td>4</td>
<td>BYU</td>
<td>Interior Design</td>
<td>U</td>
<td>F</td>
<td>French</td>
<td>Tahiti</td>
<td>5.00</td>
<td>0.50</td>
</tr>
</tbody>
</table>

*U = undergraduate; G = Graduate. ** M = Male; F = Female.

Table 3.2: Description of Course-Specific Interview Subjects

Subjects in these sections of American Heritage met for 50 minutes four days a week (Mondays, Wednesdays, Thursdays, and Fridays). Three class periods were lectures, and one was a lab directed by a teaching assistant. Both sections of this course were taught by the same professor and teaching assistant. Four tests were given throughout the semester at BYU's Testing Center, which allowed students to work at their own pace.

The course was taught using a relational learning process (see Appendix A). This methodology entails finding patterns in historical facts which can be unified under a general principle (e.g., power struggles may result in cooperation, or survival may depend on flexibility). This general principle is then connected to at least three other contexts. One context is always a student's personal life, and the other contexts can include any subject area, such as nature, biology, politics, geography, philosophy, business, or education.

Another unique characteristic of this methodology is that students are asked to draw models for homework and on tests to represent the general principles they are learning about.

According to the professor, this methodology is new to most students; therefore, students typically struggle with the course and testing expectations for the first few weeks of the semester.
Instrumentation

In order to answer the research questions considered in this study, several instruments were used, modified, and designed. In this section, the semi-structured interview, language proficiency instruments, state-anxiety measure, and the questionnaire (with background information, the testing preference scale, and the trait-anxiety scale) are described. Finally, the content-area tests are described.

Interviewing Protocol.

Interviews are purposeful conversations between two people, used for gathering descriptive information, in the interviewee's own words, about a person's opinions, attitudes, or behaviors (Bogdan & Biklen, 1992; Patton, 1990; Pedhazur & Schmelkin, 1991). Depending on research aims, the amount of structure in an interview may vary. According to qualitative research standards, when a subject is not able to "tell his or her story personally in his or her own words, the interview falls out of the qualitative range," (Bogdan & Biklen, 1992, p. 97).

Interviewing allows access to a wide variety of information quickly, with the possibility of follow-up and clarification (Marshall & Rossman, 1995). Semi-structured or open-ended standardized interviews, in particular, allow for acquiring "comparable data across subjects," (Bogdan & Biklen, 1992, p. 97). The interviewing protocol developed for the present study begins with basic background information (i.e., country of origin, length of U.S. stay, university experience, and major), and then focuses on open-ended questions about students' experience, concerns, and preferences regarding content-area testing. These questions were divided into primary and secondary concerns. Patton's (1990) content, wording, and question form guidelines were considered in developing the semi-structured interview protocol. Appendix B contains the semi-structured interview questions used in Stage One of data collection.
The semi-structured interview questions were also adapted for use in Stage Four of the study to interview the four randomly selected subjects in the classroom portion of this study. These follow-up interviews focused on subjects' testing experiences in American Heritage 100; therefore, the semi-structured questions were narrowed in focus to subjects' present classroom experience, rather than focusing on the more general experience elicited in Stage One of the study.

**Language Proficiency.**

For the purposes of the present study, language proficiency was determined using two tests: the Michigan Test of English Proficiency (MTELP), and an impromptu writing assignment. Appendix C and D have example test items from each test. These two language tests required one hour and 45 minutes: 75 minutes of out-of-class time, and 30 minutes in-class time to complete. These language tests yielded writing, reading, grammar, and vocabulary scores, which were used to explore the role of language proficiency in content-area testing, and the role of each language skill in explaining performance on various test formats (i.e., restricted response, multiple choice, and pattern model).

The MTELP (English Language Institute, 1977) is a 100 item, 75-minute, standardized multiple-choice test of grammar (40 items in dialogue format), vocabulary (40 synonym or completion items), and reading (four reading passages followed by five multiple-choice items). The test is considered appropriate for assessing the English language readiness of adult, non-native speakers of English interested in university-level academic work.

Form R of the MTELP, which was used in the present study, was tested for reliability with 772 applicants to universities. Using Kuder-Richardson Formula 21, reliability was shown to be .89, (M = 75.26, SD = 13.03, SEM = 4.13). The content, construct, predictive (GPA and classroom grades), and concurrent validities of the MTELP have also been adequately addressed and established in the English Language Institute's MTELP.
Examiner's Manual (1977). The correlation between the TOEFL (Form MEF1 or K-NEF1) and a previous version of the MTELP (Form C) was reported to be .881.

The MTELP was administered through BYU's Testing Center during the first two weeks of the semester. Scoring was done locally using a punched stencil.

The MTELP is a dated and unsecured test in comparison to current forms of the Michigan English Language Assessment Battery (MELAB) and TOEFL. Nevertheless, the MTELP offers several advantages over the use of the MELAB or TOEFL in the context of the present study. First, the MTELP could be administered, and scored locally. Second, the MTELP could be placed at a testing center, which allowed students to schedule taking of the exam at their own convenience. This minimized the demands made on students in this research context, where participation was voluntary. Third, the cost associated with purchasing the MTELP was a fraction of the cost associated with using the MELAB or TOEFL. Fourth, this study required language proficiency scores for the same semester test performance was considered; therefore, TOEFL scores at the time of admissions to BYU would not have been sufficiently current. Taking the MTELP was also a new experience for these BYU ESL/LM students, and would not be as repetitive as taking the TOEFL exam again. Fifth, the MTELP is considered a valuable predictor of performance on the MELAB and TOEFL exams. Appendix C provides examples of test items from the MTELP exam.

The impromptu writing assessment was designed to assess students general writing ability. This assessment allowed students 30 minutes of in-class writing. Although raters were not officially trained MELAB evaluators, the four OSU raters used were experienced raters, who regularly use the MELAB 10-level rating scale (English Language Institute, 1995-1996) in assigning OSU ESL students to writing courses. Appendix D contains the directions, prompts, and the scoring criteria for each of the 10 scale levels used in the rating process. Inter-rater reliability was high among raters at .91 using a Pearson Product Moment Correlation Coefficient (PPMr).

52
State-anxiety Measure.

The most common type of affective measure is the self-report instrument, which is simple to administer and score (Popham, 1988). Jones, Madsen, and Brown (1980) investigated affect in testing situations using the Jones-Madsen Affect Questionnaire. Factor analysis of this instrument revealed two main factors: emotive reactions and cognitive reactions to testing (see also Scott, 1986). Emotive reactions, for example, dealt with how pleasant, frustrating, or easy a student evaluated a test. Cognitive reactions, however, focused on issues of reliability, how well the test reflected ability, or related to course content. Madsen and Murray (1984) developed and used a state-anxiety measure based on the emotive scale of the Jones-Madsen Affect Questionnaire.

The state-anxiety measure used by Madsen and Murray (1984) was used in the present study to investigate the role of state-anxiety in content-area test performance for ESL/LM students. Following each of the four content-area tests scheduled in the two American Heritage sections, subjects were asked to complete a state-anxiety measure, and rate their preparation for and performance on the test. Appendix E contains the four state-anxiety measures used to accompany each administration of a content-area test throughout the semester.

The state-anxiety measure asked subjects to evaluate each section of the test (i.e., restricted response, multiple choice, and pattern model) by responding to three statements: 1) I liked this part; 2) This part was difficult; and 3) I felt pleasant (happy, calm) during this part. Responses on state-anxiety questions were made on a five-point Likert-type scale ranging from strongly agree to strongly disagree. Statements 1 and 3 were scored from 1 to 5, and Statement 2 was scored 5 to 1. The higher the score, the higher the level of state-anxiety.

In rating their own test preparation and performance, subjects were be asked to rate their knowledge of the U.S. testing system, and how effective their preparation for the test was
on a seven-point scale. Students also estimated their grade on the exam, identified which test format allowed them to best show their content knowledge, briefly described their preparation strategies, and noted if they planned to prepare any differently for future exams. These preparation and performance data augment descriptive analyses.

The Questionnaire.

The questionnaire, which was administered in class at the beginning of the semester, was designed to gather data in three areas: personal characteristics, testing preferences, and trait-anxiety. Appendix F contains the complete questionnaire.

Part A of the questionnaire asked for background information about a subject's gender, language background, length of residency, length of U.S. university experience, familiarity with U.S. examination system, number of previous American history courses taken, and language skills. These data contributed to investigating the role of personal characteristics on content-area test performance by ESL/LM students.

Part B of the questionnaire focused on subjects' preferences in testing, and allowed investigation of the role of preferences in relation to test performance. Zeidner's (1987) Test Attitude Inventory, used in two studies with junior high school students in Israel, was adapted to the purposes of this study. Zeidner's study compared multiple-choice and essay formats using both Likert-type and Osgood's semantic differential rating scales to obtain convergent validity on the inventory. The alpha reliability estimates for each type of scale were considered adequate for group comparisons at approximately .85.

For the purposes of this study, Zeidner's semantic differential approach was adopted, and expanded to five test formats: true/false, short answer/completion, multiple choice, restricted response, and extended essays. Subjects were asked to respond along a seven-point continuum, anchored by the following 10 adjective pairs: difficult/easy, complicated/simple, unclear/clear, boring/interesting, tricky/straightforward, unfair/fair, worthless/valuable, low expectancy of success/high expectancy of success, high anxiety
producing/low anxiety producing, and feeling uncomfortable with exam/feeling comfortable with exam. Pedhazur and Schmelkin (1991) note that this type of scale allows investigation of both directionality (difficult/easy) and intensity (how difficult/how easy). In addition, subjects’ ratings are investigated from two perspectives: relations among adjective pairs within a test format, or relations across test formats (Pedhazur & Schmelkin, 1991). The higher subjects’ ratings, the more favorable the test format is perceived.

In addition to having subjects rate test formats, subjects were also asked if their attitudes toward each of the five test formats were the same in English and their native language. If their opinion of a test format differed between their two languages, subjects were asked to briefly explain why they believed this difference existed in L1 and L2 testing situations.

Part C of the questionnaire is a measure of trait-anxiety. It is based on the Alpert and Haber (1960) Achievement Anxiety Test (AAT), which is specifically designed for academic examinations (Wine, 1980). This instrument is composed of 19 statements: 10 items measuring debilitative (items 1, 3, 4, 5, 7, 11, 13, 14, 17, 19), and 9 items measuring facilitative (items 2, 6, 8, 9, 10, 12, 15, 16, 18) test anxiety (see Appendix E). Examples of debilitative statements include "Nervousness while taking an exam or test hinders me from doing well," and "The more important the examination, the less well I seem to do." Facilitative statements include "I look forward to exams," and "Nervousness while taking a test helps me do better." Each statement is responded to on a five-point Likert-type scale ranging from rarely to almost always.

Based on Alpert’s and Haber’s (1960) original four samples, the debilitative scale of the AAT has a mean of 26.33, and a standard deviation of 5.33. The facilitative scale has a mean of 27.28, and a standard deviation of 4.27. Using the test-retest method over a ten-week period, Alpert and Haber (1960) report AAT reliability at .87 for the debilitative items, and .83 for the facilitative items. The two components of the scale are also
negatively correlated ($r = -0.37$). According to Sapp (1993), the validity and reliability of the AAT measure is well established in test anxiety research.

**Content-area Tests.**

The four American Heritage tests given during the semester were not manipulated in any way by the researcher. Each test had three sections: restricted-response, multiple-choice, and subject-pattern-model. This format was standard. Restricted-response items asked students to list, define, give reasons, or illustrate various principles/concepts from American history. The multiple-choice section, with approximately 30 items, focused readings from the textbook. Finally, the subject-pattern-model section of the test began with a reading (from as few as 280 words to as many as 1000 words) on such topics as mercantile trade, *The Declaration of Independence*, or *The Gettysburg Address*. Based on the reading provided, students were to complete a subject pattern model, which required them to write a specific principle, a general principle, give an example of the principle, and then apply the principle symbolically to four different contexts (e.g., nature, United States history, personal life, and another context of the student's choosing). Appendix G contains 1) a copy of a practice exam handout provided to students before the first exam; and 2) a copy of a subject pattern model homework sheet. No examples of multiple-choice questions are provided. Because test performance is considered holistically and according to section format (e.g., restricted-response, multiple-choice, pattern-model), test scores were also recorded according to test format.

**Pilot Study**

To field test the instruments related to the independent variables, a pilot study was carried out at Snow College in Ephriam, Utah. A five-week summer block Sociology 101 course was selected for the pilot study based on its high enrollment of ESL/LM students. Ten of the seventeen students enrolled were ESL/LM students. Students who participated in the study received extra credit. Although these ESL/LM students were matriculated, it
should be noted that their English language proficiency was lower than what would be required for study at four-year universities, such as BYU.

During the five-week course, students took four exams: two take-home exams, and two testing center exams. The testing center exams had 30 multiple-choice questions, four restricted-response questions (i.e., define), and one extended-essay question. The take-home exams required students to answer five restricted-response questions, and two extended-essay exams. The professor offered take-home exams expressly to help the ESL/LM students in his class.

The pilot study provided useful insights for improving the BYU study. For example, the ESL/LM students at Snow College were confused by the inclusion of sample test items on the questionnaire, used to define each test format. Greater clarity was achieved in defining test formats by providing only descriptive definitions on the questionnaire, and by providing examples separately on an overhead while students completed the questionnaire. The pilot study also demonstrated that extra credit should be given expressly for taking the language proficiency exams to improve participation: Only four of the ESL/LM students took the MTELP (Form F), and writing assessment. These pilot study insights resulted in the full participation of all 38 ESL/LM students in the BYU study.

Data Collection Procedures

As described previously, data collection took place in four stages. First, semi-structured, face-to-face interviews were conducted with 13 university-level ESL/LM students. Second, subjects in American Heritage 100 completed the questionnaire and impromptu writing assessment in class during the first week of the semester. Third, scores from the MTELP administered in the testing center were gathered by the end of the second week of classes. In addition, content-area test scores and state-anxiety data were gathered throughout the semester. Fourth, follow-up interviews were conducted with four
American Heritage subjects to further contextualize students' testing experiences in American Heritage 100. Figure 3.1 represents the data collection process by approach, aim, and method.

Figure 3.1: The data collection process.
Data Analysis

Data analysis was divided into a series of 5 steps, progressing from qualitative to quantitative analyses of the data. Step One entailed transcribing the tape-recorded interviews, checking transcriptions for accuracy, and conducting a content analysis of the interview data across subjects. Content analysis consists of descriptive data reduction, that is identification, coding, and categorization of significant patterns of response across ESL/LM subjects (Patton, 1990). Hyperqual 2 (Version 1.2) computer software was used to facilitate analysis (Padilla, 1993). The semi-structured interview questions served as the analytical framework for this content analysis, and was applied to both interview contexts (i.e., the 13 general interviews, and the 4 course-specific interviews). Chapter 4 contains the description of subjects' opinions, concerns, and preferences in testing, and subsequent interpretation, explanations, and conclusions emerging from the general interviews. The four American Heritage students' feedback is reported at the beginning of Chapter 5 to appropriately contextualize the quantitative findings which follow.

Verification of this qualitative analysis was addressed through the issues of trustworthiness and authenticity (Creswell, 1994). First, the triangulation of interview and questionnaire data enhanced generalizability of the data, and, therefore, the trustworthiness of the data (Marshall & Rossman, 1995; Patton, 1990). Second, the limitations of the data collection procedures are recognized, in that time spent interviewing and building relationships with interviewees was limited (Glesne & Peshkin, 1992). Third, the audio tapes and computer transcriptions of each interview have been preserved as an indication of authenticity (Guba & Lincoln, 1989).

Chapter 5 addresses the results of the course-specific data analyses, which cover steps two to five below, and answer research questions two through twelve. Step Two provided the descriptive analysis of all independent and dependent variables; namely, frequencies, means, standard deviations. Step Three consisted of plotting the data on scattergrams to
check for linear relations among variables, and influential outliers (e.g., using Cook's D). Although test anxiety generally has a curvilinear relationship with test performance (and would require data transformation or polynomial regression procedures), this was not found to be the case in the present study. Also, zero-order correlations (PPM_r) between variables were generated, answering research questions 2, 5, 6, 7, 8, 10, and 11.

Step Four of data analysis entailed multiple regression analysis. According to criteria discussed by Pedhazur and Schmelkin (1991), regression analysis is justified in the context of the present study for several reasons: 1) several independent variables—language proficiency, test anxiety, and testing preferences—were selected as potentially important variables in predicting performance on content-area tests; 2) if independent variables are only nominal, then analysis of variance (ANOVA) and regression are equally valid statistical alternatives; however, when independent variables are continuous, regression must be used (i.e., ANOVA requires dichotomizing continuous independent variables); 3) regression is "equally applicable to experimental and nonexperimental designs," (Pedhazur & Schmelkin, 1991); 4) nonexperimental researchers are more aware of correlations among variables when regression is employed; and 5) simple correlations only give a general picture of the relationship among variables, but a regression equation allows for greater specificity.

Given the focus on test-taker variables in a classroom testing context, the goal of these multiple regression analyses was prediction within the sample. As such, regression coefficients, assessment of multicollinearity, and identification of influential outliers were important. When independent variables were categorical (i.e., gender and language group), dummy-coding was used to appropriately transform the data. The criterion in this study is content-area test performance in American Heritage 100 (i.e., the dependent variable).
Research questions 3 and 4 focus on language skills and overall test performance or test performance by question format. To date current theory does not suggest that language skill variables (i.e., reading, writing, grammar and vocabulary) be entered into the regression equation in any particular order. Therefore, stepwise regression (examining semi-partial correlations and multiple R-squared) was employed.

For regression analyses carried out in Step Four, and subsequently in Step Five, it is important to underscore that this study focuses on prediction or criterion-validation in a sample. A concern for statistical assumptions for inference to a population is not required. Although the application of regression analyses is the same in experimental and nonexperimental designs, it is important to note that nonexperimental designs require a different interpretation of these analyses: Interpretation, therefore, is limited to prediction not explanation, and to the sample not the population of ESL/LM students.

Step Five of data analysis involved the application of path analysis techniques. Typically, path analysis is a multivariate approach to analyzing recursive or unidirectional causal relationships among variables, which are depicted graphically as a path diagram (Hatch & Lazaraton, 1991; Kerlinger & Pedhazur, 1976; Vogt, 1993). Path analysis is theory-driven, not data-driven. Kerlinger and Pedhazur (1976) explain that, "path analysis is not a method for discovering causes, but a method applied to a causal model formulated by the researcher on the basis of knowledge and theoretical considerations," (p. 305).

Using multiple regression analyses, path analysis "can determine whether or not a pattern of correlations for a set of observations is consistent with a specific theoretical formulation," (Kerlinger & Pedhazur, 1976, p. 317). In the present study, regression analyses can estimate the direct influence of the exogenous and endogenous independent variables of interest. In path analysis, the strength of relationships is expressed in terms of path coefficients. Vogt (1993) defines a path coefficient as "a numerical representation of the strength of the relations between pairs of variables in a path analysis when all the other
variables are held constant," (p. 167). A path coefficient is the standardized regression coefficient (i.e., beta weight), or if expressed in unstandardized form, it is called a "path regression coefficient," (p. 167). This path coefficient is symbolically expressed as $p$ with two subscripts. For example, $p_{21}$ means the coefficient for dependent variable #2 being regressed on the independent variable #1.

Path analysis requires more than one regression analysis. According to Kerlinger and Pedhazur (1976), "At each stage, a variable taken as dependent is regressed on the variables upon which it is assumed to depend," (p. 314). According to Hatch and Lazaraton (1991), the assumptions of path analysis and regression are the same.

Both regression and path analysis are large sample techniques, which require randomization to make causal claims. In the present study, these two assumptions of regression and path analysis are not met. Causal claims cannot be made (Hatch & Lazaraton, 1991). Given the exploratory aims of this study, and the logic of path analysis, Figure 3.1 represents the theoretical model to be tested (i.e., Research Question 12) using a three stage multiple regression procedure.

The five data analysis steps described herein allowed the 12 research questions posed to be answered. It was assumed that if the various independent test-taker variables were found to adequately predict content-area test performance, significant insights would be gained into testing from a second language perspective. These insights would benefit language and testing professionals concerned with understanding and/or accommodating ESL/LM students' test performance in content-area situations.
Figure 3.2: A path diagram of the variables to be investigated in predicting ESL/LM students' test performance in content-area classes.
CHAPTER 4

QUALITATIVE INTERVIEW RESULTS AND DISCUSSION

Introduction

In order to understand the ESL/LM student experience in content-area testing, 13 university ESL/LM students were interviewed using a semi-structured protocol (see Appendix A). The content analysis of these interviews reveals ESL/LM students' opinions, concerns, and preferences in content-area test situations (i.e., Research Question 1). These findings provide both test-taker and second language perspectives to what many teachers and SLA researchers consider the uncertainty for ESL/LM students in content-area test situations. Furthermore, these qualitative findings and implications are intended to compliment, if not corroborate, the course-specific quantitative analyses presented in Chapter 5.

Results

The interview questions provide the structure for presenting the results of the content analysis; namely, four main issues are addressed: 1) problems in testing (questions 5-8, and 13); 2) testing preferences (question 9); 3) testing strategies (question 7); and 4) helpful teacher practices (questions 10, 11, 12). Quotations from students are identified by both student and source card identification numbers assigned during HyperQual2 data analysis (i.e., [Subject number:Source Card ID]). Appendix H contains an example transcript of Hyperqual2 output for Subject 8.
Problems in Testing.

ESL/LM students were asked to respond to several questions focusing on problems in testing. For example, students were asked to describe their first experiences in testing, note American and home-country differences in testing, and identify the problems they encounter when taking content-area tests at the university level. They were also asked whether they believed their grades on tests accurately reflected their content knowledge. From these questions, these ESL/LM students revealed problems in testing related to context, culture, and language. Elaboration of these problems in testing follows.

1. The Context. The ESL/LM students interviewed were mindful that they were a minority population in their American university classrooms. These students were aware that they compete with native English speakers for grades, and 75% of these ESL/LM students felt that their grades did not accurately reflect their content knowledge and ability.

According to these ESL/LM students, grading in content-area courses was an issue where students and professors shared responsibility. For example, one student accepted that academic success depended on her own efforts: "We have to study harder... twice as hard" as native speakers [02:14284/17488]. For another student, his test performance fluctuates, and feels almost beyond his own control: "My life depends on midterms, homework due. I am absolutely unsure about myself right now. Sometimes I do good; sometimes I have to retake [a class] one more time. The first time I received a D, and then the second time I received B. I don't know." [06:36319].

Some students, however, also expressed concern that grading practices of professors played a part in ESL/LM students' grades not accurately reflecting students' content knowledge. For example, one student said poor grades were not "because we don't understand, or we don't study. We might spend twice as much time as others, but we didn't get a better grade as others," [09:53614]. Another student described a lack of fairness in grading this way:
in physics we wrote a paper. He [the professor] didn't look exactly at the content. He just looked at vocabulary words, grammar . . . [but] you have to look at what's inside—just not look at the outside. Maybe some people aren't good in it [i.e., writing]. They have the best content out of all American students, but they don't know how to write it . . . I think that's not fair. [07:42881]

These ESL/LM students also seemed to recognize the difficulty professors may face in meeting the differing needs of native and nonnative speaker in the same class. They neither expected professors to "talk in a special way for foreign people" [02:17488], or alter testing practices on their behalf. One ESL/LM student said, "I'm afraid [of tests] . . . because it's not my language. . . . teachers must take care that we are internationals . . . I don't want a different test for native speakers and international students, but like simple questions, and simple answers," [04:28154].

Still another student speaks candidly of the frustration she experiences as a student when placed in a position of training professors to work with ESL/LM students:

I try to understand why some teacher will understand me better than the other. And then you will find that this teacher never have this kind of student before. And then, you mark yourself as 'Test Student Number One.' You've got to . . . let him know what you are, what kind of problems you [have], and then he knows that he get result. . . . [who]ever comes after [me], he knows how to handle it. [09:53973]

These interview excerpts reveal that the American university context puts minority ESL/LM students in the uncomfortable position of competing with native speakers for grades, and of articulating in English their learning and testing needs to content-area faculty. These aspects of context are problematic, and contribute to the disadvantage ESL/LM students' perceive themselves as having in some content-area classes.

2. The Testing Culture. Seventy-five percent of the ESL/LM students interviewed felt that the American system of testing differed from their home-county systems. These students appeared to have experienced a gradual shift to what one subject described as "American ways," [03:20631].
Five characteristics typify the cultural differences in testing that ESL/LM students perceived between American and home-country testing systems. First, for many students, one characteristic of the American system is giving tests with greater frequency than would be expected in their home countries. Multiple midterms and comprehensive finals seem to pressure ESL/LM students: "Here in BYU I take a lot of tests for one subject, at least, three or four times a semester. That was quite a lot for me, and that is kind of a stress for me. That's why I don't like a lot of tests," [03:19464].

A second characteristic of the American system, which is new to some students, is that they are expected to know facts as well as be able to express their personal opinions. For example, one student said that in his home country, "It's more like you have to answer this because you read that... [but in an architecture class] here you have to be yourself, and you have to [write] what you think," [08:45191].

For these ESL/LM students, a third characteristic of the American system is that American professors are perceived as being more caring than expected. For example, one student described her surprise at a professor's willingness to meet her needs. This professor permitted her to use a dictionary for her second music history test following poor spelling and vocabulary use on her first exam. Because of this adjustment, she reacted by saying, "Professors... thought about me, about my problems and they... go to meet my problems, and to relief it. I got used to 'my problems are my problems. I have difficulties with English... it's my difficulties,'" [05: 30109]. Another student said, "Here, the teacher more involved with students... they explain everything," [12:63702]. Finally, another student observed that American teachers "make sure everybody understand... then give you the test," [09:50620].

For several of these ESL/LM students, a fourth characteristic of the American testing system is that it is perceived as being easier than their home-country systems. The
Russian, French, Taiwanese, and British (for an Arabic student) testing systems, for example, were each named as being more demanding than the American system.

A Taiwanese student explained this difference in testing rigor as follows:

In Taiwan, they try to make things difficult because we are small. We don't have enough space for everybody to come to college, so they make everything hard . . . to pick the best out of those people [to] go to colleges. [In America] we have a chance to come to college. We have a better education chance, and they try to make you really understand the courses. . . . I like it better here. [09:52941]

Several students also mentioned that grading in their home countries was more subjective than the American system. Students explained that in their home countries their grades are based on one final exam, often in the form of an oral exam, where professor can decide on grades subjectively: "If he likes you, it's $A$, or $A-$, or $B$. If he doesn't like you, it can be a $B-$," [01:9406].

A fifth characteristic of the American testing system centers of alternatives in testing formats which are unfamiliar to ESL/LM students. While some students were comfortable with multiple choice and essay formats; other students were not. Test format expectations were clearly dependent on each students' previous home-country experience; however, the use of a testing center to administer tests, class presentations as tests, or the idea of take-home tests were each mentioned as unique testing alternatives in the American system.

Concerning take-home tests, one student seemed surprised at the trust professors had in their students: "Such a trust to students. . . you are alone in the room. They trust you to use nothing. I tried to be honest. I used only my own brain, but it is just impossible in every other university I was studying in," [05:30109]. According to another student, the expectation of honesty takes effort because, "Cheating. . . is kind of acceptable in Korea," [03:19464].

For ESL/LM students, the differences in testing practices which characterize the American culture of testing are both positive and negative. While they struggle with the
number, format, and alternatives in testing practices, they also seem to perceive the American system as easier for them when compared to their home-country alternatives.

3. The Language. Beyond adjusting to the problems of context and culture in content-area testing, ESL/LM students also describe language-related problems that interfere in the process and products of testing. Various students mention individual problems in testing, such as finding time to study with work responsibilities, not using adequate preparation strategies, or drawing a blank when confronted with a test. More significantly, however, these ESL/LM students' responses revealed language problems in testing that were common to a majority of ESL/LM students interviewed. In order of salience in the data, these seven specifically language-related problems are 1) vocabulary, 2) understanding test questions, 3) memorizing in English, 4) knowing how to write, 5) time constraints, 6) "trapped" content knowledge, and 7) keeping pace in the classroom. Each of these issues is discussed below.

First, the most frequently identified problem in testing for ESL/LM students is a limited receptive and productive English vocabulary. From the perspective of these ESL/LM students, their failure in testing situations can be attributed, in many cases, to one word, or the one word they did not know.

Single words are first a classroom problem, where relevant test information is presented by professors to students. For example, in a linguistics class, one student explained how the entire lecture hinged on understanding what Black English was. From her perspective, as a second language learner, "It seems to me that everybody know, but I don't. And it was essential because if I know I could understand what was going on," [02:14791].

Other students said they felt lost in a flood of terminology--"big words"-- in such classes as accounting, anthropology, geography, or computer programming. One student said, "The teacher was like speaking another language. I was like 'Wow! What is he saying?' I was too lost, even though I read before the lesson in the book," [04:27631].
When ESL/LM students discussed their vocabulary knowledge and its connection to testing, they spoke in terms of lacks, limitations, and need for improvement. One student said, "My vocabulary is too short. I need to improve it. I need to write the same word in different way," [04:25771]. Another student said, "Sometimes I can't find the right word I'm thinking," [07: 40299]. Another student worried that in writing test answers, "If you don't know some word, you have to change the whole sentence because it doesn't make sense. You lose a lot of time," [08:47178].

Several comments by these ESL/LM students illustrate this point: demonstrating content knowledge on tests can be hampered by not recognizing or being able to produce one important word. For example, one student described this one-word phenomenon as follows: "Prevail--I don't know what prevail is, still don't know. That was in my midterm. I know the answer... but they word them out like college words... so I don't know if I know the answer," [10:56238]. Another student reinforced this idea that a single vocabulary item can stop an ESL/LM student from even attempting to demonstrate their content knowledge:

I don't have this big vocabulary, so I'm limited to... certain amount of words I know. If there comes a question with maybe a word that means the same thing, but I don't know that word, even when I actually know the answer... I just can't answer the question because I don't know what that word means, [11:59826].

These ESL/LM students seem to concur with this student's sentiments: "If I don't know [that word], maybe I'm going to answer something wrong." [08:45605].

The second language-related problem ESL/LM students have in testing situations is understanding the questions they are asked on tests. One student explains, that his main concern on every test is "Will I understand what the professor is asking? Most of the time, that's the problem I have. I really don't understand what the professor is asking for," [11:58740/59826].
Several students mentioned that they often misunderstand questions on tests because of how a professor chooses to write the questions. These ESL/LM students said they reread questions, analyzed questions word by word, and struggle with difficult vocabulary in order to understand test questions. One student explained that in classes such as history or chemistry, she specifically tries to remember from lectures and readings "how they ask those questions," [09:50821].

A third problem in testing for ESL/LM students is the amount of memorization required for most tests. For these ESL/LM students, memorizing content for tests is complicated by the fact that they are memorizing in their second language. One student said, "It is hard [to] memorize in Portuguese, imagine in English," [02:14284]. Other students described memorizing content-area information as a "double effort," [04:27783], or "harder" because it is "my second language," [06:35766]. One student even suggested that her ability to memorize is different in her two languages: "In Korea, I think I really good at memorizing, and somehow in USA, I forget how to memorize," [12:64013].

A fourth problem in testing for ESL/LM students is knowing how to express their knowledge in writing. Several students observed that they knew what they wanted to say, but just did not know how to write it, or express it in an appropriate "western" style [13:68487]. One student said she used to just speak "Russian by English words," and that she has had to struggle to learn to use "English rules in English structures in English essays," [05:30256]. ESL/LM students also seem to worry that some professors want them to write "like native American people," and this expectation is unfair: "They can't expect you have perfect grammar," [08:48507].

Time is the fifth problem that jeopardizes ESL/LM students' performance on tests. Specifically, these students felt that time restraints led to both language and content errors in testing. Students pointed out that in writing on tests, they "have to think, to organize, and to check grammar," [08:45605], and that their ideas often come to them first in their
native language, and need to be rephrased in English. These language-related processes require time. One student's comment typifies this ESL/LM language concern, complicated by time in testing situations, as follows:

"Time is sometimes a factor in my performance on an exam. Honestly, it takes me a little bit more time to read and actually understand clearly... Most of the time, I know the answer, [but] I do something wrong because I did it really quick... sometimes I would like to have more time [11:61934]

A sixth problem that these ESL/LM students described in testing was having content knowledge "trapped" in a portion of their mind, or in their native language in such a way that they could not adequately access that knowledge to demonstrate mastery in test situations. For many ESL/LM students this phenomenon is described as knowing some content or word in their native language, but not knowing it in English. For one student taking a religion class, she said, "I could remember the scripture in Portuguese, but I could not translate," [02:14284]. A Korean student described taking an anthropology class when he first came to the United States, because he had already taken a similar class in his native language. He had hoped that this background knowledge would give him an advantage; nevertheless, he reported that, "I know the answer of the question, but the problem is I cannot describe in English," [03:19804].

Finally, the seventh problem mentioned frequently by ESL/LM students was their difficulty in keeping pace from a language perspective with the teacher in the classroom. This problem was described as becoming accustomed to American expressions and accents, taking notes and listening at the same time, or producing university level work. One student described her professor's pace as follows: "One class about 200 people... he doesn't know who has the questions or not. He write fast. Talk fast. [You] write down everything, and you have to go home and figure out the rest of it," [09:53973].

From the ESL/LM student perspective, these seven language-related problems in content-area testing are unique to second language students, and jeopardize their ability to
demonstrate content-area knowledge on tests. The context, culture, and language problems for ESL/LM students in content-area testing can be further contextualized by understanding ESL/LM students' testing preferences, testing strategies, and ideas regarding helpful teacher practices. A discussion of these issues follows.

Testing Preferences.

When these ESL/LM students were asked about their testing preferences in content-area courses, they responded in general, and were then specifically asked about five format alternatives: true/false, multiple choice, short answer/completion, restricted response, and essays. Appendix I contains examples of each format alternative referred to during interviewing. Students' general and specific preferences in testing are described in turn.

When ESL/LM students were asked about their general testing preferences, two answers were common. First, several students noted that they preferred oral tests because that was what they were accustomed to in their home-country experience. One student observed that if the professor didn't understand him, "I can try to explain... we can debate: what you think, or what I think. Why it is right, or why it is wrong," [08:46718].

Second, students felt that their format preferences would depend on the subject-matter of the course. For example, in a religion class, one student would prefer to "do a lot of writing," [04:19928]. For physics or math, one student preferred multiple-choice tests because "you can take out [answers] that don't make sense... so that helps you out," [07:41787]. Another student felt that it is better to be tested "chapter by chapter" in a math class. For another student, test preferences depend on whether the course content is technical or theoretical. He felt that multiple-choice formats were more appropriate for "technical" material, and essays would be more suitable for "theory" courses, such as political science [11:60137].

When asked about the five specific test format preferences (e.g., true/false, multiple choice, etc.), ESL/LM students preferred multiple-choice test formats most often. They
also felt that multiple choice tests were the easiest for them to do well on in testing situations. Essay exams and short answer/completion test formats were considered the most difficult by these ESL/LM students. When asked if these format preferences would be the same in both their native and second languages, these students responded either that 1) all formats would be easier in their native languages, or that 2) their preferences would be the same for both languages.

When asked why certain test formats were or were not preferred, these students had very little to say about true/false, or short answer/completion test formats. Generally, they reported that true/false items were confusing, tricky, or picky, but provided a fifty percent chance of getting a correct answer. Short-answer/completion items were disliked because the format required memorizing a specific word. No positive comments were offered by these ESL/LM students regarding short-answer/completion questions.

Multiple-choice items were preferred by these ESL/LM students, and generated commentary from 12 of the 13 students. Although considered the easiest format, students offered strong criticisms of the format as well. For example, students felt that multiple choice questions were tricky, ambiguous, and difficult to understand when time was limited. They also felt that some professors gave too many options to choose among. The main strength of multiple choice items, according to these ESL/LM students, is that the correct answer is in front of them. As one student said, "If I know, I can figure it out. We have all answers. If I know materials, I can figure it out." [01:11035].

Most students felt positive about restricted response test items, which require them to briefly list, define, or give reasons. While two students felt it was "difficult to explain something very short," [05:32200], other students felt that it was a good opportunity to give "simple definitions or simple answer[s]," [03:21470], or "gain a lot of points on test[s]," by writing, at least, something down [09:53973].
Finally, ESL/LM students were the most negative about essay-format tests. Although essay questions were credited for allowing a student to "write down every single thing" s/he knows [11:60341]. Students also agreed with the sentiment of the following student regarding essay questions: "My native language is not English, so to write extended essay will be tough on me... when you tell me to use the limit[ed] amount of words I have learned in English, and try to make a coherent academic essay, it will be quite tough." [13:70913]. These students mentioned grammar, punctuation, syntax, vocabulary, the challenge of developing a topic, and difficulty in expressing their own opinions as some of the reasons they did not like essay questions on exams.

Overall, these ESL/LM students showed strong preference for multiple-choice questions, which rely more on reading skills than writing skills. Restricted-response and essay items, however, were still preferred over short answer/completion items. From these students' feedback, it appears that they have strong opinions concerning which test formats help and hinder their test performance in content-area classes. Although test-taker characteristics, such as gender or language background, did not emerge as clear aspects of preference, it is possible that familiarity with the American testing system is associated to ESL/LM students' preferences in content-area testing.

**Testing Strategies.**

ESL/LM students, like all students, approach testing situations idiosyncratically. Strategies which work for one student, may not fit with another student's study habits. Nevertheless, several of the ESL/LM students interviewed appeared to relate to the following student's struggle in content-area classes:

It's hard to memorize for me. I can read this, and I can understand, but if you make me a question of this, I can't do it. I don't know how to explain it, but I can't take a test based on what I read... I read like [I'm reading] a magazine, not like I need to study. Sometimes there are few words I can't understand. Is it important? And I don't know how to take notes. I'm just copying what the teacher [does] in the blackboard," [04:27783].

75
This student identifies his lack of study skills, specifically academic reading, note-taking, and vocabulary skills, as obstacles to his academic success.

Table 4.1 lists various strategies these ESL/LM students mentioned related to coping with tests in content-area courses. Broadly, these strategies address what students considered important in preparing for a test, or while taking tests. The test-preparation strategies discussed reflect three components; namely, 1) principles that holistically guide students' study efforts; 2) use of various resources; and 3) specific study techniques.

In terms of guiding study principles, students seemed to recognize the need to adjust study skills to the demands of different classes:

Each class has a different format. One of the key components is the syllabus. . . . I read the syllabus carefully and prepare for the first test. After the first test, I know what will be there in the second exam or what kind it will be, [03:20631].

These ESL/LM students also acknowledged that each class, subject-area, and department has a structure that needs to be mastered for academic success. For one student, these structures guide his study priorities:

If I go to business school, they gonna watch particular classes, especially math. So I'm gonna put some pressure on me on math class. I'm not going to worry about American Heritage and English 115. I know at the same time that GPA is really important if I want to get scholarship or something, [01:10299].

In terms of resources in the test preparation process, ESL/LM students recognized professors, teaching assistants, course syllabi, study guides, textbooks, and classmates as important sources of information. Connecting effectively with these resources is, at times, challenging for ESL/LM students. One young lady, who is fluent in speech, complained that because of her oral language skills:

Those teachers won't believe me when I say I have a problem: I can't understand them. My English TA once say, 'I don't think you [have] any problems. As long as you read the books, you come to see me, you'll be okay.' I have problems understanding the literature. When I ask for help, you tell me I'll be fine. It's not the answer I'm looking for here, [09:53973].
Preparation-related strategies:

- **Guiding Principles**
  "Know the structure of the subjects"
  "Carefully prepare for the first test... will know what will be in the second exam"
  "I study hard at the beginning of the semester to get a good grade in first test"
  "You have to know everything"

- **Resources**
  "The main thing is the teacher"
  "I can break down questions with my classmate"
  "I am asking my friends in English if I have a doubt: what does this word mean"
  "TAs are most helpful... they tell you what the professor is looking for"
  "The key components is the syllabus"
  "I like study guide"
  "Get the practice exam from the instructor as fast as possible"
  "Never ignore the book even if the professor ignores it"

- **Specific Techniques**
  "I mark my books"
  "Take a lot of notes"
  "I always summarize"
  "If something I don't know, I'd leave it to the very end [to read] before the exam"
  "Reading a lot because when you read you know the order of the sentence"
  "I will try to pick up the most important point in the test... and memorize"
  "I try to connect things... I make a picture... I try to connect with my real life"
  "I just memorize key word and generate"
  "I try to remember everything as a story [for essay exams]"
  "Now I try to ask more questions if I don't understand"
  "I prefer to listen to the professor, ask questions, and forget about taking notes"

Test-related strategies:

"First I look through the exam. If something easy... I concentrate on [that]"
"I first read it, and make sure what it's about"
"I will skip the hardest one first if I find I am stuck"
"I just keep going. If I have time, I go back to the question I still have to answer"
"I read the question, I think the answer, and if the answer is in the paper, it's that"
"I try to find if they have some question in the next page he gives the answer"
"If you cannot find the answers [on multiple choice], write down a B or C"
"Write anything down, just relate to the question... they will give you points"
"I mostly will take a best guess"
"The dictionary helps me"
"I like to pay attention in everything--word by word"
"Ask my instructor to explain"
"Before I write the essay, I will use the brainstorming technique"
"They give you a scratch paper, and I just write ideas so I won't forget"

Table 4.1: ESL/LM Student Strategies for Coping in Content-Area Tests

77
Despite the challenges associated with getting the help they need, several ESL/LM students noted that these people or paper resources help them sort through what content is important for a particular test. This sorting process in test preparation is pivotal: "You have to memorize... everything... It's kind of hard because he [the professor] put a lot of stuff that you don't even think was important," [08:45605].

ESL/LM students also described various strategies they employ while being tested in content-area courses. Many strategies mentioned were idiosyncratic. For example, one student, in particular, is trying to find study strategies that work:

I have some difficulty with the language, so if I try to take notes and listen, I will end up not understanding anything. I was trying to do two things at the same time. Now, I prefer to listen to the professor, ask questions, and forget about taking notes. That's one thing I had changed... it seems to be working pretty well so far. [11:58968]

These ESL/LM students also had strategies they used that were common among them. For example, skipping difficult questions, reading questions carefully, and using dictionaries typify these more common strategies. Although students' comments revealed awareness of test-taking strategies, it is unclear how effectively these strategies are applied to content-area test situations by ESL/LM students.

Helpful Teacher Practices.

Students were asked to indicate what teaching practices their professors could use that would help them perform better on content-area tests. These requests of content-area teachers are summarized in Table 4.2 as classroom and test preparation requests. These requests reflect many of the issues students' addressed as testing problems and strategies. For example, one student's comment underscores the importance of vocabulary selection and clear question writing as helpful teacher practices for the ESL/LM population: "The same test form, but use different words. Try to make the question clear, same answers, same question to make people like us really understand the questions... take out all those big words," [09:53369]. Overall, the helpful teacher practices these ESL/LM students
In-Class Requests:
Don't speak so fast
Explain difficult concepts twice
Reinforce what is and is not important in the textbook
Provide study guides
Discuss the format of upcoming tests
Tell us specifics for how to improve test performance
Be available during office hours
Call failing students in for a conference
Allow for extra credit
Have smaller classes

Test Preparation Requests:
Write test questions clearly
Use simple vocabulary
Allow more time during tests
Allow dictionary use during tests
Have some available to explain difficult test questions

Table 4.2: ESL/LM Students Requests of Content-Area Teachers

identified show a concern for 1) access to useful test-preparation resources; and 2) professors' language use in test writing.

Finally, these ESL/LM students were also asked to reflect on which language skill—reading, writing, listening, speaking, grammar, vocabulary, or pronunciation— was most important in helping them show their content knowledge on tests. They responded with the greatest frequency that improving their reading skills would be most helpful to their academic success. One student explained that "Each class I take they have books with the class. I need to understand what the books say. I need to read the books very fast before the midterm or the final. . . . That's one of the main input for me," [13:71676]. The second most frequently mentioned language skill for helping students demonstrate their knowledge on tests was the writing skill, followed in turn by vocabulary and grammar.

Commenting on the role of ESL programs and teachers in preparing students for testing in university courses, one student said: "I think they [i.e., ESL teachers] didn't prepare me
for the university. They just prepared me to pass the TOEFL, and that's it." [04:27783].

Although this was only one student's opinion, other students mentioned the need for ESL programs to focus on such topics as university expectations, writing timed essay responses, increasing the reading component of ESL classes, and addressing study skills (e.g., note-taking and memorization strategies).

Through these interview data, these students have provided an insider's view of the problems, preferences, strategies, and preferred teacher practices associated with content-area testing for ESL/LM students. Their responses demonstrate the importance of context, testing culture, and language in understanding content-area testing from a second language perspective; however, they also suggest that format, strategies, and teacher practices potentially impact ESL/LM student performance in various content-area test situations. The implications of these findings are discussed in the following section.

Discussion

The purpose of the interview portion of this study was to illustrate qualitatively the potential problems in testing that ESL/LM students experience in content-area courses; namely, to explore the opinions, concerns, and preferences of ESL/LM students that suggest associations between various test-taker variables and test performance in content-area courses (see Figure 2.1).

The purpose of the discussion which follows is to situate the qualitative findings of this study in the context of current theory and research. The relevance of these findings to SLA, psychology, and psychometrics are addressed by focusing on how language proficiency, test anxiety, and testing preferences relate to content-area test performance for ESL/LM test-takers.

1. Language Proficiency. Two findings stand out regarding the role of language proficiency in content-area testing. First, ESL/LM students describe their language-related ESL/LM problems in content-area testing in terms of their academic language skills, or what
Cummins (1980) termed cognitive/academic language proficiency (CALP). This distinction between interpersonal and academic language proficiency is well-supported in the second language literature (e.g., Cummins, 1980; Kinsella, 1992; Snow & Brinton, 1988). Furthermore, Cummins (1980) argues that academic language abilities are interdependent in learners' first and second languages. Although these students have obtained a level of second language proficiency that gained them university entrance, their comments reveal that they do not assess themselves to be adequately proficient in academic language use—below a threshold of adequate academic language proficiency—to accurately or consistently demonstrate their content knowledge on tests.

Second, these students identified reading and writing as the macro-skills which would help them most in testing; however, they spoke most frequently about how their limited vocabulary jeopardized their test performance in content-area classes. Through their examples and elaborations, these ESL/LM students intimated that vocabulary knowledge is the sub-skill of greatest importance in testing. Leki and Carson (1994) also argue that vocabulary is key in academic achievement. This, of course, does not deny the role of vocabulary in reading, writing, listening, or speaking in the academic context. These findings do, however, lend tentative support to Matthews' (1990) argument that the sub-skills of language proficiency not be awarded equal weights in the proficiency equation.

2. Test Anxiety. Although test anxiety is an empirically measured psychological construct, the ESL/LM students interviewed provided evidence of both the emotional (test anxiety), and cognitive (test wisdom) aspects of test taking (Spielberger & Vagg, 1995b). For example, ESL/LM students' test anxiety was rooted in such language problems as their inability to 1) recognize, produce, or unleash the right words; 2) understand test questions; or 3) memorize content in English. In addition, these students reacted emotionally to various test formats: Several students had aversions to one particular test format or another (e.g., multiple choice or essays). These students were also anxious over competing with
native speakers, and the grading practices of some professors. Finally, test wisdom was also revealed in these students' use of guiding principles, resources, and specific strategies in preparing for, and taking content-area tests. Students' expectations in testing were clearly culturally framed as suggested by such researchers as Scott (1986), and Diaz-Guerrero (1976).

These findings qualitatively corroborate what the psychology literature predicts concerning test-anxiety for minority populations. Perceived poor performance, insecurity, and fear of negative evaluation—which were each expressed in these interviews—are evidence of test anxiety (Anderson & Sauser, 1995). Research has also demonstrated that worry—negative thinking—is the cognitive aspect of state-anxiety which is most strongly associated with poor test performance (e.g., Sapp, 1993). Differences in test anxiety related to gender, language background, or socio-economic status were not detected through this interview format; nevertheless, these data point to aspects of language, culture, and testing preferences as potential explanations of test anxiety among ESL/LM students.


These findings support previous descriptive work related to ESL/LM students' preferences in testing. Qualitative studies by Leki and Carson (1994), Horowitz (1986), Kinsella (1992), and Leki (1995) describe writing as a problem for items that require production, and time limits as a problem for objective items. A quantitative pilot study by Madsen and Murray (1984) also suggest that ESL/LM students prefer objective test formats.
The testing preferences of ESL/LM students also appear to be similar to native speakers of English. Native speakers show preferences for restricted-response and essay items over multiple-choice formats on the criteria of "value, fairness, and validity in assessing content knowledge," (Zeidner, 1987, p. 357); however, they believe multiple choice tests are easier to do well on (e.g., Gellman & Berkowitz, 1993; Zeidner, 1987).

This preference for multiple-choice type tests by ESL/LM students in content-area courses draws into question criticisms of standardized, national exams which are identified as lacking reliability and validity for minority and language-minority students (e.g., Anderson, 1988; Lam, 1993; Lemke, 1990; Madaus, 1994; Soto, 1992). Although ESL/LM learners have language, anxiety, and format obstacles to overcome in testing, it is uncertain if multiple-choice test formats create more or less problems in testing for ESL/LM students than other format alternatives. It is also unclear whether format preferences would vary in classroom and national testing contexts.

General Implications

These opinions, concerns, and preferences reveal how ESL/LM students orient to testing in content-area courses. Furthermore, these qualitative findings improve our understanding of the problems ESL/LM test-takers experience in expressing content knowledge in a second language. The implications for theory, classroom practices, and research of these qualitative findings are discussed below.

First, the qualitative findings suggest that the disadvantages in testing for ESL/LM students should be theoretically extended to explicitly recognize both internal and external dimensions. According to these data, the internal dimension for ESL/LM students is comprised of linguistic, cognitive, and affective aspects. Test performance for ESL/LM students appears to be influenced by their language proficiency, in particular their ability to comprehend language used in the classroom, and on tests. The linguistic aspect includes mastery of the grammatical, sociolinguistic, rhetorical, and strategic in language
development. For example, poor reading and writing skills may hinder a student from responding or comprehending a test question correctly. On the other hand, if a student lacks strategic competence, s/he may not be able to compensate for a limited vocabulary.

Cognitive aspects which may influence test performance include intelligence, memory, analysis, and metacognitive strategies. ESL/LM students may face obstacles in test situations when their cognitive processing is impeded by difficulties in memorizing, accessing content knowledge, or applying appropriate test-taking strategies.

The affective aspect of testing for ESL/LM students involves such things as attitudes, emotions, motivation, and personality. The findings indicate that ESL/LM students experience some level of anxiety (i.e., worry) in content-area test situations, and that this anxiety may be heightened or alleviated by students' testing preferences, or cultural expectations.

The internal dimension of a ESL/LM student's disadvantage in testing is within the learner's development and control. The external dimension in testing is outside the ESL/LM students' control. For example, a student's performance on a test may be influenced by the vocabulary, test format, time constraints, or grading practices a professor chooses to use. The pedagogical practices of teachers may also support or hinder student learning of the content knowledge required for tests.

These internal and external dimensions of being tested in a second language clarify and define the nature of content-area testing for ESL/LM students from a qualitative perspective. Whether students' cognitive or linguistic aspects of the internal dimension are more important than affective aspects is undetermined. Similarly, whether the internal outweighs the external in influencing ESL/LM students' testing outcomes is unclear. Understanding the nature of ESL/LM students' potential disadvantage in expressing content
knowledge, however, can improve how the test performance of ESL/LM students is interpreted, and facilitated by content-area teachers.

Second, the qualitative findings described suggest ways in which to extend, build upon, and adjust current educational practices to meet the needs of ESL/LM students in content-area classes. For example, both intensive and matriculated ESL programs should continue to tailor program curricula to meet the academic needs of university-bound students, incorporating more explicit attention to test taking issues. ESL/LM students should be provided instruction that helps them develop the linguistic, cognitive, and affective abilities that would minimize the obstacles in content-area testing within their control. In addition, ESL teachers need to be equipped with skills that make them powerful advocates for ESL/LM student populations in diverse content-area settings.

Third, the qualitative findings suggest research into the role of vocabulary knowledge in content-area test performance be explored more fully. More research is also needed to clarify how expressing content knowledge through various test formats influences test performance. Multiple-choice, restricted-response, and essay formats appear to be the most important formats to distinguish among. In addition, students' preference for multiple-choice test formats goes against what advocates for ESL/LM populations claim regarding the bias in standardized national tests. These issues in content-area testing should be incorporated into future research designs.

Limitations

The findings and implications of these interview data reveal the opinions, concerns, and preferences of ESL/LM students in content-area testing situations. These qualitative results delineate the problems ESL/LM students experience when expressing their knowledge in English. Furthermore, these findings provide an insider's view of content-area testing.

Qualitative research does not pretend to be replicable (Marshall and Rossman, 1995), rather it reveals the complexity of the testing context for ESL/LM students. Because
subjects were small in number, not selected by random sampling procedures, limited to the university-level, and interviewed with a semi-structured protocol, the results of this qualitative study must be interpreted cautiously. Claims of causality or generalization to other populations cannot be made. Nevertheless, these qualitative findings are supported by, and augment available anecdotal, case study, and practitioner articles (e.g., Leki, 1995; Rosenthal, 1996; Short, 1994). Within these descriptive limits, the findings can be said to illuminate the issues for ESL/LM students in content-area testing.

Future Directions

Efforts to investigate content-area testing for ESL/LM students from a qualitative perspective provides a vehicle through which ESL/LM students' voices, as test-takers, can be heard. As Silverman (1993) notes, field research is validated by "comparing different kinds of data" and by using "different methods... to see if they corroborate one another," (p. 156). Research into the testing concerns of ESL/LM students in content-area classes is scarce and exploratory. There is clearly a need for further research, both qualitative and quantitative, to clarify the relationships among various test-taker variables, such as language proficiency, test anxiety, and format preferences.

This qualitative study intimates that research with primary and secondary populations regarding content-area testing could be valuable. In addition, research addressing the sub-skills of language proficiency, the memorization task from a second language perspective, and format preferences according to subject-areas would be informative.

Conclusion

The qualitative portion of this study provides a gestalt of the concerns, problems, and preferences ESL/LM students have when demonstrating their content knowledge on tests. In order to explore the quantitative influence of language proficiency, test anxiety, and testing preferences on content-area test performance, correlational research was carried out in a university-level, American Heritage course. The findings and implications of this
quantitative research is presented in Chapter 5. The qualitative and quantitative findings of this study will be synthesized in Chapter 6 to present implications for ESL programs, ESL teacher preparation programs, and future collaboration with content-area faculty.
CHAPTER 5

COURSE-SPECIFIC RESULTS AND DISCUSSION

Introduction

As described in Chapter Three, data were collected from 38 subjects in an American Heritage course to investigate content-area testing from a second language perspective; in particular, to examine the obstacles that ESL/LM students may face in testing. The follow results answer the research questions related to exploring associations among test-taker variables (i.e., language proficiency, test anxiety, and testing preferences), and test performance in a university-level history class, which has been tailored to the needs of international students.

The purpose of this chapter is four-fold. First, using qualitative interview and course evaluation data from students, the American Heritage course is described to appropriately contextualize the present study. Second, quantitative results related to subjects' test performance, language proficiency, test anxiety, and testing preferences are presented. Third, results related to a path analysis of the theoretical model formulated for this study (see Figure 2.1 & 3.1) are addressed. Fourth, a discussion of these findings is presented, with attention to the implications, limitations, and future directions of second language content-area testing research.

Qualitative Results

In order to understand and describe the ESL/LM student experience in two international sections of American Heritage, four students were randomly selected to participate in end-of-semester interviews about their experience. In addition, course evaluation data were
gathered by the professor about the effectiveness of the Relational Learning Approach employed (see Chapter 3 or Appendix A). The results of these data address Research Question 1 and 2 from a qualitative, course-specific perspective of 38 ESL/LM subjects.

The Interviews.

The four students selected for end-of-semester interviews were asked to evaluate their experience in American Heritage by responding to the qualitative interview protocol (see Appendix B). For consistency, the content analysis of students' responses are presented in the framework employed in Chapter 4; that is, results focus on problems, preferences, strategies in testing, and helpful teacher practices.

1. Problems in Testing. To uncover the problems ESL/LM students may have had in their American Heritage class, students were asked to describe their first testing experience in American Heritage, identify any problems they had, and reflect on whether their test grades accurately represented their content knowledge. The findings which follow demonstrate that these ESL/LM students perceived few problems in the American Heritage testing context.

A. The context. ESL/LM students in these international sections of American Heritage were not a minority, but rather in the majority. Although 23 of their classmates were native English speakers, these students were, nevertheless, international students. None of the four students interviewed mentioned having to compete with native English speakers in this academic setting. Indeed, three out of the four students interviewed felt that their test grades in American Heritage accurately reflected their content knowledge.

According to these ESL/LM students, the use of a relational learning process methodology was unique to both their home country and United States university experiences. The relational learning methodology relied on deriving general principles from the history content to be learned. These general principles were represented graphically, as models, in lectures, and applied not only to history, but to the students'
lives, and other content-areas (see Appendix A). On tests students were asked to draw and explain their own models of various general principles. Although the learning approach was new to these ESL/LM students, they viewed it positively. For example, one student commented that, "the model part is very interesting... I like it because it helped me to understand many things... The models can be applied to anything," [02:11657].

B. The testing culture. Five differences in the testing practices in American Heritage were identified by these ESL/LM students as being different from their home-country experiences. First, each student mentioned that the test was long, and took more time than they anticipated. For example, students took as little as 90 minutes to as long as 5 hours to complete the first test. One student observed, that the professor "did all that he needed to for that exam. He covered everything," [01:8181].

Second, one student felt that her home country's system of testing was harder: "French education is harder than here," [04:18754]. Third, however, another student described the American system as more objective than her home country system:

This written test are more fair, I mean because they are the same for everybody... it's standard. [In my home country] if he [the professor] likes you, he can give you some easy question, and if he doesn't like you, he can told you some very bad questions, and you'll never pass," [03:15317].

Fourth, these students also identified differences in test formats between the United States and their home countries. Two students felt that in their home countries there would be greater reliance on essay and oral test formats over multiple choice formats. Another unique aspect of the American Heritage course was the requirement to illustrate models on tests (e.g., see Appendix G). Students noted that illustrating content-area concepts was new to them: "For other tests if we don't know the concepts, we don't answer. In this test, we have to think about the model," [02:11496]. He continued saying, "It was different because we have to draw, and we have to understand the models... [on] other tests we have to know, [but] in this one we have to think," [02:11496].
Finally, two students mentioned that they were not comfortable taking tests in the testing center. Each of the four course exams were administered in a testing center, where students could take the exams at their convenience over a two day period. None of the tests had a time limit, and students were allowed to use their dictionaries. Nevertheless, one student discovered that the testing center made her nervous. Another student observed: "I don't like the testing center. There are too many people there. I prefer just a classroom, with my classmates." [02:13535]. Overall, out of these five observations, only the length of the tests, and the use of a testing center were considered negative in the American Heritage testing culture.

C. Other Concerns. Contrary to the findings mentioned in Chapter 4, none of the four American Heritage students mentioned any specifically language-related problems. The problems they described center more in adjusting to the test format, and in preparing appropriately for tests.

For the first exam, two students mentioned that they were confused about what the professor was looking for. One student said, "At first when I open the test, I didn't know what he wanted... but then I just start to write everything I know," [01:8181]. For another student, the multiple-choice section of the test, which was based on the readings, was difficult. According to the students, the professor did not highlight the important parts of the text in class; therefore, two students underestimated the text's importance while studying, choosing to focus more on class notes. For example, one test covered ten chapters in the text, and a student said, "I didn't really know what to focus on," [03:15078].

Each student described memorization strategies used in preparing for the American Heritage tests; however, their comments also reveal memorization as a test preparation concern. For example, one student was anxious about memorizing for the final comprehensive exam:
I know memory is fragile. I'm not sure I'm gonna be able to memorize all the things that I was supposed to memorize for all these three tests. . . . Because you need to put every single word and every single arrow or star. . . . It takes me like two days to memorize all this stuff. [01:8181/8799].

2. Testing Preferences. Each of the four American Heritage exams were similar in format, comprised of three sections: restricted-response, multiple-choice, and subject-pattern-model questions (see Appendix G or Chapter 3). Given this context, students were asked which section of these tests was the easiest and most difficult, and which allowed them to best demonstrate their content knowledge.

All four students interviewed felt that the subject pattern model was the easiest section on the test, because they had similar weekly homework assignments. One student said, "That was okay because at that point we had been doing some of those models for homework. I knew what he was looking for." [03:15078]. Another student noted that everything she needed to answer was given: "I mean the text is there, read it. You just have to understand it, and get a pattern in there, and apply it. . . . This class trained me for that," [04:19928].

Three of the four students felt that the most difficult part of the test was completing the restricted-response section. One student thought the multiple-choice section was the most difficult. All four students, however, identified the restricted-response section, which require them to write and draw, as the section which allowed them to best show their mastery of content knowledge.

Students said that the restricted-response section was the most difficult because they had to prove what they knew. This section took them more time to complete, and several mentioned that having to write and draw was new to them. One student observed about the restricted-response section of the test: "You have to think; you have to remember; and you have to express in the written, and through pictures," [03:16307]. Another student,
however, noted that, "If you don't know how to express your ideas, at least, you can
draw something, and you can get some points out of that," [01:8462].

There was less consensus among the four students regarding the benefits or drawbacks of the multiple-choice section of the test. One student did not like multiple-choice questions on any test, while the other students liked it for various reasons. For example, one student liked it because "multiple choices you can answer without reading," [01:8462]. Another student mentioned that because "we have choices maybe we can get it right even though we don't know it," [02:12687]. Finally, a third student said, "Because you are exposed with all the answers, you don't have to search your brain to find the answer. From there, it's mainly what you learn--it's systematic. You can just use logic," [04:19928].

One common complaint among these students about the multiple-choice section of the test was that students felt the professor did not adequately emphasize what was or was not important in the readings. Although materials for lectures were supported through examples in the text, students felt they needed more help in knowing what was important to study for the tests. Finally, each of these students felt their testing preferences in this American Heritage course would be the same whether the tests were taken in English or their native languages.

3. Testing Strategies. The ESL/LM students interviewed mentioned preparation-related strategies more than test-taking strategies. Among the preparation strategies mentioned, students highlighted three principles which guided their preparation; namely, attending class, reading the text, and knowing the pattern models. These students also saw classmates and study groups as an important resource in preparing for their tests.

The specific test-preparation techniques students described using for their American Heritage class focused on memorization strategies. Students said they began memorizing early for the exams, because there was too much material to memorize in a day or two. One student described her memorization process as follows:
I have this white board at home, so I just quiz myself. I said 'draw this model,' and I would draw the model myself at home. Sometimes I start memorizing a week ahead. I have to go back three times to get it in my brain. First time I just read it. The second time I will try to memorize it a little more... [then] I have to write it down. [04:19222].

Only one student mentioned a specific strategy used while taking a test. He noted that "I first go for the easy ones... the ones I know, and then I go for the hard ones," [02:12135]. Nevertheless, several students mentioned while discussing the multiple-choice section of these tests, that they used the options provided to guess or figure out the correct answer. Overall, students' answers revealed that their best test-taking strategy for success was based on how effectively they could memorize the pattern models, with their accompanying labels and arrows.

4. Helpful Teacher Practices. These four American Heritage ESL/LM students were very positive about both their professor's and teaching assistant's teaching practices. Their professor was described as using clear and simple language in lectures, and on tests. One student explained as follows:

Some other professors I have use tricky questions. So sometimes I don't really know that word that they are using. . . . but this one [the professor] is very clear. I mean he asks us for very specific things. . . . He uses a clear language, and we can understand it easily. [02:12948].

Several students noted that the professor's practice of writing everything on the blackboard was particularly helpful to them. For example, one student said, "He writes everything for us--every single thing that we need to write," [01:9950]. Another student observed that, "The way that he presented the material was very clear. . . I never felt I needed to go to his office, and ask for additional clarification," [03:16627]. This same student goes on to explain, "He never used difficult language, or if he used some words that weren't very common, he would just use a synonymy, so we have the hard word, but he'd [use] also bunch of other easier words." [03:16825].
The American Heritage professor also helped his ESL/LM students by providing a practice test prior to the first exam, by giving no time limit, by not focusing on grammar mistakes, and by allowing students to use dictionaries. One student described the benefit of dictionaries as follows:

My problem is that I get stuck on a word, and I can't finish the sentence. I have to be sure about the spelling, and if I can just look it up, I will be much fine with that, [03:15608].

These students were also very positive about their teaching assistant, who helped them during office hours, and was also willing to attended their study group sessions. The only type of help students needed, which they felt was lacking, was guidance in handling the reading assignments for class. One student felt that the teaching assistant needed to "tell us what is really important for us to learn in the book. Because he just say 'Chapter 25;' but you don't know which kind of questions you're gonna get," [01:9273]. Another student was frustrated by the text, and said, "A lot of times I didn't understand the text... it was frustrating to go and read it, and reread it again, asking friends. That was a pain," [04:20130].

When asked how their American Heritage course compared to other courses they were taking, students' comments suggest that they do experience some problems as ESL/LM students in other classes. For example, one student explained as follows:

I don't know the words in computer science. There's nobody who can explain. There are many students who have a lot of harder questions for the professor... It is supposed if I am there, I know how to speak, so they don't really get it from any of us international students on that, [02:13076].

Another student observed that vocabulary is also a problem for her in photography and interior design classes. She said:

I was lacking a lot of vocabulary... It's kind of frustrating when you don't understand one word, and she goes on using that word again, and again. 'What is it we're talking about?' Yeah, I didn't have to feel clueless in his class [i.e., American Heritage], [04:20474].
Course Evaluations.

The American Heritage professor asked students to rate how effectively various components of the course helped them understand and apply the relational learning process. Table 5.1 lists these components in rank order by average student rating, which was based on a 10-point scale, where 10 was the highest rating. Student evaluations show that the teacher, class presentations, relational models, and teaching assistant were the most effective tools in helping students to understand the relational learning process. Course exams were considered the least effective in reinforcing relational learning processes.

<table>
<thead>
<tr>
<th>Component</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Teacher</td>
<td>9.50</td>
<td>0.81</td>
</tr>
<tr>
<td>The Class Presentations</td>
<td>8.85</td>
<td>1.46</td>
</tr>
<tr>
<td>The Relational Models</td>
<td>8.67</td>
<td>1.41</td>
</tr>
<tr>
<td>Teaching Assistant</td>
<td>8.50</td>
<td>1.72</td>
</tr>
<tr>
<td>The Semester Project</td>
<td>8.39</td>
<td>1.61</td>
</tr>
<tr>
<td>The Homework</td>
<td>8.36</td>
<td>1.76</td>
</tr>
<tr>
<td>Classmates</td>
<td>7.39</td>
<td>2.17</td>
</tr>
<tr>
<td>The Exams</td>
<td>7.00</td>
<td>1.84</td>
</tr>
</tbody>
</table>

Table 5.1: Mean Ratings of How Effectively Course Components Reinforced Relational Learning
Qualitative Discussion

These qualitative results are important in placing the American Heritage course, and its testing protocol, into a context leading to appropriate interpretation of students' content-area test performance. In particular, seven findings from this course-specific, qualitative data emerge as important in shedding light on the problems, preferences, strategies, and teacher practices in testing which characterize content-area testing for ESL/LM students (i.e., Research Questions 1 and 2).

Three findings are important to understanding problems in testing. First, in this American Heritage setting, ESL/LM students constituted a majority, and these students felt that their test grades accurately reflected their content knowledge. Second, although test length, format, and use of a testing center differed from these students' home-country experiences, these students felt that the American culture of testing was easier, but also more fair, than their home-country systems. Third, students identified two problems in their American Heritage testing experience; namely, difficulties sorting out what was important to learn from the course textbook, and employing effective memorization strategies.

In the American Heritage testing context, problems in testing differed from those described in general interview findings presented in Chapter 4. These American Heritage students did not mention specifically language-related problems, such as vocabulary, confusing questions, trouble writing, time limitations, trapped content knowledge, or keeping pace. They did, however, mention the intimidating task of memorizing large amounts of materials, and comprehending the class text as being problematic.

Two findings clarify students' course-specific preferences in testing. First, these ESL/LM students identified the restricted-response section of the test as the most difficult section, but also the section which allowed them to best demonstrate their content knowledge. Second, multiple-choice questions were considered helpful to ESL/LM students because students had the answers in front of them, and could guess correctly.
using logic. This view of multiple-choice questions is supported in native speaker research by, for example, Gellman and Berkowitz (1993), Nield and Wintre (1986), and Zeidner (1987).

These course-specific testing preferences are consistent with the findings discussed in Chapter 4. Although students may prefer formats which are perceived as easiest for them (subject pattern or multiple choice), these students lend support to the idea that what is most difficult (i.e., restricted-response questions in this case) is also the test format that allows them to best demonstrate their content knowledge. Which test formats give students the most points, and which test formats allow them to demonstrate their knowledge best appear to be separate issues.

One finding stands out in regard to students' use of strategies for test preparation and test-taking: these ESL/LM students appear to have strategies in place, on which they rely, to help them succeed in content-area testing settings. Although these four students did not describe a broad variety of strategies, their responses suggest the importance of memory and memorization as important aspects of the second language testing experience.

Regarding helpful teaching practices, ESL/LM students' comments suggest that their professor minimized the obstacles in testing for them by employing specific practices. In contrast to the challenges students described in second language content-area testing in Chapter 4, these American Heritage students appear to have ideal conditions for success in testing. Not only were they a majority population, these American Heritage ESL/LM students praised their teacher for clear language use in class and on tests, had access to dictionaries in testing, could illustrate as well as write their knowledge, and took tests without time constraints. Students' final grades were also based on multiple indicators: four exams, weekly subject pattern model assignments, and a class presentation.

This professor's approach to teaching American history is tailored to meeting the needs of ESL/LM students, and supported by SLA pedagogical literature. For example, Short
(1994) described successful social studies instruction as meeting 12 criteria: 1) gives opportunities to communicate in oral, written, physical, or pictorial forms; 2) connects content taught to real-life experiences of students; 3) taps students' native country knowledge; 4) activates background knowledge; 5) provides hands-on or performance-based activities; 6) promotes critical thinking; 7) makes language accommodations that help students learn; 8) uses graphic organizers; 9) seeks cooperative learning and peer tutors; 10) provides models; 11) opens discussion to different perspectives of history; and 12) adjusts instruction for different learning styles. Students comments provide evidence that this American Heritage context was pedagogically organized to benefit ESL/LM students.

When these ESL/LM students compare their American Heritage experience to other classes, they suggest that in other classes they do experience language-related problems, specifically vocabulary limitations. Findings in Chapter 4 and SLA literature (e.g., Leki & Carson, 1994) support this concern over vocabulary for ESL/LM students in most content-area settings. Nevertheless, that vocabulary was not a concern for students in the American Heritage context is noteworthy.

Conclusion

Overall, the four ESL/LM students interviewed felt that the professor, teaching assistant, lectures, and test format were very suitable and helpful in meeting the need of ESL/LM students in content-area classes. These ESL/LM students felt the course made them think, and was particularly useful because course content was applied to their personal lives. According to these ESL/LM students, they experienced relatively little, if any, problems in demonstrating their content knowledge on tests. This, therefore, validates the qualities of instruction and testing which appear to minimize the second language jeopardy in content-area testing from a qualitative perspective.

The quantitative results and discussion which follow focus on exploring further how, and if, test-taker variables, such as language proficiency, test anxiety, or testing
preferences, impact content-area test performance for ESL/LM university students. These findings remains peculiar to the American Heritage course experiences of 38 ESL/LM students.
Quantitative Results

In order to quantitatively investigate content-area testing from a second language perspective, questionnaire and test performance data were gathered from ESL/LM students in two sections of American Heritage for international students. In this section of the findings, descriptive test performance data (i.e., the dependent variable) is presented first in order to frame subsequent discussion of ESL/LM students' content-area test performance. Test performance data is followed by findings related to each of the endogenous independent variables: language proficiency, test anxiety, and testing preferences. The findings related to the theoretical model, which unites exogenous and endogenous independent variables, are then presented. The results which follow answer research questions 2 through 12.

Test Performance.

The ESL/LM students in American Heritage took four exams during Fall semester, 1996. Each exam was administered in a testing center, where students were allowed as much time as needed, and the use of dictionaries. In addition, each test began with a write and draw or restricted-response section, followed by multiple-choice and pattern-model sections (see Appendix G). Table 5.2 contains means and standard deviations of students' test performance by section (i.e., test format) on the four content-area tests.

Concerning students' test performance, three findings are noteworthy. First, students overall mean performance improved slightly, showing less variability, throughout the semester. Second, students obtained the most points on each test from the pattern-model section, with very little variability among students. Third, students did better overall, but also showed more variability in performance, on the restricted-response sections than multiple-choice sections. This pattern only varied on Test 3, where students' mean performance was better on the multiple-choice section than on the restricted-response section.
<table>
<thead>
<tr>
<th>Format</th>
<th>Test 1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted Response</td>
<td>M 75.94</td>
<td>77.92</td>
<td>77.66</td>
<td>82.94</td>
<td>78.63</td>
</tr>
<tr>
<td></td>
<td>SD 17.75</td>
<td>15.40</td>
<td>14.10</td>
<td>12.12</td>
<td>12.30</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>M 71.88</td>
<td>76.03</td>
<td>78.22</td>
<td>65.11</td>
<td>72.95</td>
</tr>
<tr>
<td></td>
<td>SD 10.49</td>
<td>11.77</td>
<td>13.66</td>
<td>10.19</td>
<td>8.52</td>
</tr>
<tr>
<td>Pattern Model</td>
<td>M 98.09</td>
<td>98.75</td>
<td>99.28</td>
<td>98.92</td>
<td>98.76</td>
</tr>
<tr>
<td></td>
<td>SD 4.63</td>
<td>3.89</td>
<td>1.41</td>
<td>2.54</td>
<td>2.21</td>
</tr>
<tr>
<td>Overall</td>
<td>M 78.74</td>
<td>81.15</td>
<td>81.23</td>
<td>81.33</td>
<td>80.67</td>
</tr>
<tr>
<td></td>
<td>SD 10.62</td>
<td>10.35</td>
<td>10.36</td>
<td>8.55</td>
<td>8.42</td>
</tr>
<tr>
<td></td>
<td>N 38</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 5.2: Mean Scores of Students' Test Performance by Format

Table 5.3 presents the frequency and percent of students' test-format preferences for best demonstrating content knowledge on American Heritage tests. Results indicate that students felt they demonstrated their content knowledge best on the restricted-response section of the tests, followed in turn by the pattern-model, and multiple-choice sections. This pattern was only altered for the comprehensive final (i.e., Test 4), where a majority of the students felt that the pattern-model section best demonstrated their knowledge, and restricted-response and multiple-choice sections were equally preferred by the remaining students.

Three times during the semester (i.e., the first week of classes, after the first exam, and after the final exam), students were asked to rate how knowledgeable they considered themselves to be about university-level testing in the American system, using a seven-point
Table 5.3: Test Format Preferred for Best Demonstrating Content Knowledge on Tests

<table>
<thead>
<tr>
<th>Format</th>
<th>Test 1</th>
<th>Test 2</th>
<th>Test 3</th>
<th>Test 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restricted Response</td>
<td>20</td>
<td>17</td>
<td>16</td>
<td>10</td>
<td>41.4</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>6</td>
<td>14</td>
<td>10</td>
<td>10</td>
<td>26.3</td>
</tr>
<tr>
<td>Pattern Model</td>
<td>11</td>
<td>6</td>
<td>10</td>
<td>17</td>
<td>28.9</td>
</tr>
</tbody>
</table>

N 37 37 36 37

scale. Students' ratings showed U-shaped tendencies regarding how knowledgeable they considered themselves to be about testing; that is, following the first exam, students considered themselves less knowledgeable (M = 4.28, SD = 1.24) about the American system of testing than they had at the beginning of the term (M = 4.41, SD = 1.11); however, at the end of the term, they rated themselves as more knowledgeable (M = 4.74; SD = 1.17) than they had the first week of classes.

As part of evaluation completed after their first test, students were asked to describe how they studied for the first exam, and whether they would prepare differently for subsequent tests, having seen the type of exam their professor would give throughout the semester. The most common test-preparation strategies students listed were 1) reading the text; 2) reviewing lecture notes; 3) studying with friends or in a small group; 4) studying models and diagrams; and 5) trying to memorize. Individual students mentioned the teaching assistant, reviewing the practice test, and anticipating questions as other test-preparation strategies.
Out of the 36 students who provided feedback on their first testing experience in American Heritage, 31 students (86%) said they would prepare differently for subsequent exams. The most common tactic mentioned for preparing for future tests was to study everything in more depth. For example, one student wrote on his evaluation, "I wouldn't skip anything. Everything is on the test." Other students mentioned improving their memorization strategies, spending more time reading the textbook, studying the models, and improving their note-taking efforts as future test-preparation strategies they would consider.

Language Proficiency.

In order to gather current data related to students' English language proficiency, subjects took the Michigan Test of Language Proficiency (MTELP, Form R), and completed a 30-minute impromptu writing assignment. In addition to these quantitative measure of language proficiency, students were asked to identify their strongest language skill, and which language skill they felt would help them most when taking content-area exams. These findings address research questions 2 through 4.

Table 5.4 summarizes students' mean test performance on the writing assessment, and MTELP. Students performed best on the grammar and writing assessments, followed in

<table>
<thead>
<tr>
<th>Skills</th>
<th>M</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total MTELP Score</td>
<td>76.70</td>
<td>12.38</td>
<td>37</td>
</tr>
<tr>
<td>Grammar</td>
<td>82.36</td>
<td>14.24</td>
<td>37</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>73.06</td>
<td>13.36</td>
<td>37</td>
</tr>
<tr>
<td>Reading</td>
<td>73.65</td>
<td>18.88</td>
<td>37</td>
</tr>
<tr>
<td>Writing Score</td>
<td>82.28</td>
<td>6.02</td>
<td>38</td>
</tr>
<tr>
<td>Overall Composite</td>
<td>77.79</td>
<td>10.29</td>
<td>37</td>
</tr>
</tbody>
</table>

Table 5.4: Mean Language Proficiency Scores by Skill Area

104
turn by the reading and vocabulary section. These ESL/LM students' average score on the MTELPP is roughly equivalent to scoring between 500 and 550 on the TOEFL (English Language Institute, 1977).

Table 5.5 presents students' responses to which micro- or macro-language skills they considered their strongest. Seventy-four percent of these ESL/LM students identified the receptive language skills of listening and reading as their strongest. None of the 38 students identified writing as their strongest language skill.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Listening</td>
<td>16</td>
<td>42.1</td>
</tr>
<tr>
<td>Reading</td>
<td>12</td>
<td>31.6</td>
</tr>
<tr>
<td>Speaking</td>
<td>8</td>
<td>21.1</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>1</td>
<td>2.6</td>
</tr>
<tr>
<td>Grammar</td>
<td>1</td>
<td>2.6</td>
</tr>
<tr>
<td>Writing</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 5.5: Students' Perception of Their Strongest Language Skill

Results presented in Table 5.6 show that students reported that reading was the most important language skill in helping them on tests at the university level. Writing and vocabulary skills were also considered important in testing, whereas grammar was not selected by any student as important to testing.

<table>
<thead>
<tr>
<th>Skill</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>15</td>
<td>39.5</td>
</tr>
<tr>
<td>Writing</td>
<td>10</td>
<td>26.3</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>9</td>
<td>23.7</td>
</tr>
<tr>
<td>Speaking</td>
<td>3</td>
<td>7.9</td>
</tr>
<tr>
<td>Listening</td>
<td>1</td>
<td>2.6</td>
</tr>
<tr>
<td>Grammar</td>
<td>0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Table 5.6: The Language Skill That Students Feel Helps Most When Taking Tests
For research questions 2, 3 and 4, analyses showed that the correlation between English language proficiency (MTELP) and test performance was not significant, nor did any of the language skills of reading, writing, grammar, or vocabulary predict ESL/LM students' overall or format-specific performance in this American Heritage context. Nevertheless, students indicate that reading, writing, and vocabulary are important language skills in testing.

Test Anxiety.

Test anxiety was investigated using both trait- and state-anxiety measures. Alpert and Haber's (1960) Achievement Anxiety Test (AAT), which specifically measures academic test anxiety, was used to gather trait-anxiety information. State-anxiety measures were administered following each of the four content-area exams to measure students' situation-specific level of anxiety (see Appendix E). The trait- and state-anxiety findings which follow address research questions 4 through 7.

1. Trait-Anxiety. Table 5.7 presents means and standard deviations of students' responses to the 10 statements which comprise the debilitative anxiety scale, and the 9 statements which comprise the facilitative anxiety scale of the trait-anxiety measure. Responses were given on a five-point scale, where 1 meant rarely, and 5 meant almost always.

Students reported having roughly equal amounts of debilitative ($M = 31.29, SD = 6.23$) and facilitative ($M = 30.18, SD = 4.75$) test anxiety, with an overall mean of 62.05 ($SD = 9.13$). The correlation between the debilitative and facilitative scales indicated a positive, moderate correlation between the two scales, $r (36) = .55, p < .001$. Thirty percent of the variance in these two scales is accounted for by this association ($r^2 = .30$).

To determine if these sample means and correlations significantly differed from the native speaker population studied by Alpert and Haber (1960), $t$-tests were carried out. Results showed that ESL/LM students' debilitative mean score was significantly higher.
Debilitating Items  | M  | SD  
--- | --- | ---  
1. Nervousness while taking an exam or test hinders me from doing well. | 3.08 | 1.12  
3. In a course where I have been doing poorly my fear of a bad grade cuts down on my efficiency. | 3.52 | 1.03  
4. When I am poorly prepared for an exam or test, I get upset, and do less well than even my restricted knowledge should allow. | 3.32 | 1.09  
5. The more important the examination, the less well I seem to do. | 2.63 | 1.02  
7. During exams or tests, I block on questions to which I know the answers, even though I might remember them as soon as the exam is over. | 2.95 | 1.11  
11. I find that my mind goes blank at the beginning of an exam, and it takes me a few minutes before I can function. | 3.00 | 1.27  
13. I am so tired from worrying about an exam, that I find I almost don't care how well I do by the time I start the test. | 2.79 | 1.17  
14. Time pressure on an exam causes me to do worse than the rest of the group under similar conditions. | 3.12 | 1.21  
17. I find myself reading exam questions without understanding them, and I must go back over them so that they will make sense. | 3.79 | 1.07  
19. When I don't do well on a difficult item at the beginning of an exam, it tends to upset me so that I block on even easy questions later on. | 3.03 | 1.03  

(Table 5.7, continued) 

Table 5.7: Summary Statistics for the Debilitative and Facilitative Scales of Trait-Anxiety
Facilitating Items | M     | SD
---|---|---
2. I work most effectively under pressure, as when the task is very important. | 2.68 | 1.27
6. While I may (or may not) be nervous before taking an exam, once I start, I seem to forget to be nervous. | 2.53 | 1.15
8. Nervousness while taking a test helps me do better. | 4.13 | 1.12
9. When I start a test, nothing is able to distract me. | 3.13 | 1.14
10. In courses in which the total grade is based on one exam, I seem to do better than other people. | 3.68 | 1.04
12. I look forward to exams. | 4.11 | 1.20
15. Although "cramming" under pre-examination tension is not effective for most people, I find that if the need arises, I can learn material immediately before an exam, even under considerable pressure, and successfully retain it to use on the exam. | 3.03 | 1.13
16. I enjoy taking a difficult exam more than an easy one. | 4.05 | 0.98
18. The more important the exam or test, the better I seem to do. | 2.84 | 0.89

Note. N = 38.

than language majority population, t (37) = 4.91, p < .05. Similarly, ESL/LM students' facilitative anxiety was significantly higher than the language majority population, t (37) = 3.77, p < .05. In addition, the sample correlation (r = .55) between the debilitative and facilitative scales was significantly higher than the language majority correlation (r (377) = -.37), t (36) = 3.95, p < .05.

The correlation between trait-anxiety and overall test performance (i.e., Research Question 4) was negative, and statistically significant, r (36) = -.60, p < .001. Thirty-six percent of the variance in ESL/LM students' test performance is accounted for by students' trait-anxiety (r² = .36). High levels of trait-anxiety are associated with lower test scores.
Zero-order correlations between length of residency, university experience, familiarity with the testing system, or previous American history course work were run (i.e., Research Question 5). Only students' self-assessment of their familiarity with the American testing system (M = 4.4, SD = 1.1) proved to be negatively and moderately associated with students' trait-anxiety, r (37) = -.38, p < .02. This suggests that students with higher levels of trait-anxiety rated themselves as having less knowledge of the American testing system. This correlation, however, only accounts for 14% of the variance between the two variables (r² = .14).

2. State-Anxiety. To obtain state-anxiety scores, students indicated how strongly they agreed or disagreed (using a five-point scale) with three statements: 1) I liked this part; 2) This part was difficult; and 3) I felt pleasant (happy, calm) during this part. Students responded to these statements for each test format: restricted-response, multiple-choice, and pattern-model. Adopted from Madsen's study (1981), a score of 3 on an absolute scale would suggest an anxiety allaying test format, 9 suggests an anxiety threshold, and 15 suggests a highly anxiety producing test format.

Table 5.8 presents means and standard deviations for state-anxiety for each exam by question format. Students' state-anxiety ratings revealed that the restricted-response section of these content-area tests were the most anxiety producing, reaching an anxiety threshold overall. The pattern-model section of these tests consistently were rated as causing the least state-anxiety.

In examining individual test results, students rated the multiple-choice section of the first test as more anxiety producing than either the restricted-response or pattern-model sections. Nevertheless, following the first exam, the multiple-choice questions were consistently rated as less anxiety producing than the restricted-response questions. Although state-anxiety ratings were relatively consistent across exams, students gave the third exam the highest state-anxiety rating.
Table 5.8: Mean Score State-Anxiety Ratings on Course Exams by Question Format

On average, state-anxiety scores were not significantly associated with students' overall test performance (Research Question 6); however, state-anxiety was found to negatively and moderately correlate to performance on the final exam, $r(29) = -0.42, p < .02$. The higher a student's state-anxiety, the lower his/her score on the comprehensive final. Students' state-anxiety ($M = 25.60, SD = 5.20$) accounted for only 18% of the variance in performance on the final exam.

Among the personal characteristics investigated (Research Question 7), length of residency showed a moderate, negative association with state-anxiety ($r(37) = -0.36, p < .05, r^2 = .13$), suggesting that state-anxiety decreases with length of residency. In addition, the number of American history courses students have taken previously showed a moderate, negative association with state-anxiety ($r(37) = -0.42, p < .01, r^2 = .18$).
Length of residency and previous history experience, however, account for relatively small amounts of variance in state-anxiety. The personal characteristics of length of university experience and familiarity with the testing system were not correlated to state-anxiety.

**Testing Preferences.**

The purpose of this section of results is to present ESL/LM students' general preferences in testing related to true/false, multiple-choice, short-answer/completion, restricted-response, and extended-essay formats (see questionnaire in Appendix D). Course-specific testing preferences, which were presented in the Test Performance section above, showed that in the American Heritage context ESL/LM students scored higher, and most preferred restricted-response questions for demonstrating their content knowledge.

Research questions 9 through 11 address students' general testing preferences. In order to ascertain these format preferences, students were asked to assess each format using a seven-point semantic differential scale comprised of 10 adjective pairs (e.g., difficult/easy, unclear/clear, etc.). These data provide both overall composite scores and individual scores by adjective pairs for each of the five test formats. Following presentation of composite and adjective pair findings, results are also presented related to whether students' attitudes toward each format were the same in their first (L1) and second (L2) languages.

First, students were asked to select among the five test formats which format allowed them to best demonstrate their content knowledge. Table 5.9 presents the frequency and percent of ESL/LM students' format preferences. Results show that, among the five test formats considered, approximately one-third of the ESL/LM students felt that multiple-choice tests are the best for demonstrating content knowledge; however, short-answer/completion and extended-essay test formats were also more frequently preferred formats. When test formats are considered in terms of whether they require students to write/produce answers, or read/recognize answers, a greater frequency of students selected production formats over recognition formats for demonstrating their content knowledge.
Table 5.9: Question Formats Student Feel Best Allow Them to Demonstrate Knowledge

For comparison, Table 5.10 presents means and standard deviations for composite scores derived from student ratings of each format based on the ten adjective pairs. The higher the mean score, the more favorable ESL/LM students' disposition is toward the question format. Results show that ESL/LM students were most favorably disposed to short-answer/completion items, followed by restricted-response and multiple-choice formats. ESL/LM students were least favorably disposed toward extended essays. This ranking of test formats differs from those presented in Table 5.9, when students merely selected categorically among the five formats. Nevertheless, question formats which required students to write/produce answers were generally rated more favorably by ESL/LM students than recognition question formats, which is consistent with findings presented in Table 5.9.

An analysis of variance of the results presented in Table 5.10 was carried out, using a repeated-measures approach, to determine if students' mean score ratings of the five test formats differed significantly from each other. The omnibus F-test was significant,
Table 5.10: Mean Composite Ratings of Question Formats on a Semantic Differential Scale

\[ F(41, 147) = 1.92, p < .003, \] suggesting statistically significant mean differences among the five test formats. Nevertheless, the Scheffé follow-up procedure for making multiple nonorthogonal post hoc comparisons was not significant. Because power to detect differences, and control of the experimentwise error rate are in conflict, there is not sufficient evidence in this study to pinpoint where the mean differences lie among these five test formats. An analysis of variance considering production and recognition format mean differences was not significant.

Table 5.11 summarizes ESL/LM students' rating of each test format by adjective pair. Within each adjective pair, test formats are ranked in order from least to most favorable. ESL/LM students rated true/false and extended-essay formats more often at the extreme ends of the adjective scale. For example, ESL/LM students rated true/false formats the most negatively. They felt comfortable with true/false questions, and considered true/false questions easy; however, they rated them as the most unclear, unfair, boring, worthless, and tricky format. Students also rated true/false questions as giving them a low expectancy of success.
<table>
<thead>
<tr>
<th>Bipolar Adjective Pairs</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Difficult to Easy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended Essay</td>
<td>3.27</td>
<td>1.47</td>
</tr>
<tr>
<td>Restricted Response</td>
<td>3.73</td>
<td>1.54</td>
</tr>
<tr>
<td>Short Answer/Completion</td>
<td>4.38</td>
<td>1.22</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>4.40</td>
<td>1.42</td>
</tr>
<tr>
<td>True/False</td>
<td>4.88</td>
<td>1.20</td>
</tr>
<tr>
<td><strong>B. Complicated to Simple</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extended Essay</td>
<td>3.37</td>
<td>1.43</td>
</tr>
<tr>
<td>Restricted Response</td>
<td>3.73</td>
<td>1.42</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>4.20</td>
<td>1.39</td>
</tr>
<tr>
<td>True/False</td>
<td>4.27</td>
<td>1.38</td>
</tr>
<tr>
<td>Short Answer/Completion</td>
<td>4.54</td>
<td>1.21</td>
</tr>
<tr>
<td><strong>C. Unclear to Clear</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True/False</td>
<td>4.02</td>
<td>1.28</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>4.18</td>
<td>1.57</td>
</tr>
<tr>
<td>Extended Essay</td>
<td>4.49</td>
<td>1.24</td>
</tr>
<tr>
<td>Restricted Response</td>
<td>4.58</td>
<td>1.39</td>
</tr>
<tr>
<td>Short Answer/Completion</td>
<td>4.67</td>
<td>1.38</td>
</tr>
<tr>
<td><strong>D. Boring to Interesting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>True/False</td>
<td>4.28</td>
<td>1.35</td>
</tr>
<tr>
<td>Extended Essay</td>
<td>4.42</td>
<td>1.51</td>
</tr>
<tr>
<td>Restricted Response</td>
<td>4.45</td>
<td>1.26</td>
</tr>
<tr>
<td>Short Answer/Completion</td>
<td>4.50</td>
<td>1.23</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>4.60</td>
<td>1.29</td>
</tr>
</tbody>
</table>

(Table 5.11, continued)

Table 5.11: Mean Score Ratings of Test Formats by Bipolar Adjective Pairs
(Table 5.11, continued)

<table>
<thead>
<tr>
<th>E. Tricky to Straightforward</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>True/False</td>
<td>3.04</td>
<td>1.41</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>3.26</td>
<td>1.64</td>
</tr>
<tr>
<td>Extended Essay</td>
<td>4.53</td>
<td>1.21</td>
</tr>
<tr>
<td>Short Answer/Completion</td>
<td>4.71</td>
<td>1.50</td>
</tr>
<tr>
<td>Restricted Response</td>
<td>4.74</td>
<td>1.45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>F. Unfair to Fair</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True/False</td>
<td>4.08</td>
<td>1.25</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>4.12</td>
<td>1.57</td>
</tr>
<tr>
<td>Extended Essay</td>
<td>4.30</td>
<td>1.42</td>
</tr>
<tr>
<td>Short Answer/Completion</td>
<td>4.56</td>
<td>1.59</td>
</tr>
<tr>
<td>Restricted Response</td>
<td>4.78</td>
<td>1.15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>G. Worthless to Valuable</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True/False</td>
<td>4.23</td>
<td>1.34</td>
</tr>
<tr>
<td>Multiple Choice</td>
<td>4.44</td>
<td>1.50</td>
</tr>
<tr>
<td>Short Answer/Completion</td>
<td>4.76</td>
<td>1.22</td>
</tr>
<tr>
<td>Restricted Response</td>
<td>5.08</td>
<td>1.09</td>
</tr>
<tr>
<td>Extended Essay</td>
<td>5.15</td>
<td>1.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>H. Low Expectancy of Success to High Expectancy of Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>True/False</td>
</tr>
<tr>
<td>Multiple Choice</td>
</tr>
<tr>
<td>Restricted Response</td>
</tr>
<tr>
<td>Extended Essay</td>
</tr>
<tr>
<td>Short Answer/Completion</td>
</tr>
</tbody>
</table>

(Table 5.11, continued)
The extended-essay format was also rated as an extreme for five of the adjective pairs. ESL/LM students were most uncomfortable with the extended-essay format, and rated extended essays as the most difficult, complicated, and high-anxiety producing format. On the other hand, ESL/LM students also rated essay tests as the most valuable format alternative.

Short-answer/completion, restricted-response, and multiple-choice formats were rated more moderately. Nevertheless, short-answer/completion questions were preferred for being simple, clear, and providing a high expectancy of success. Multiple-choice questions were considered interesting, and low-anxiety producing for ESL/LM students. ESL/LM students preferred restricted-response questions for being the most straightforward and fair test format.
Table 5.12 lists the means and standard deviations for production and recognition test formats by adjective pair. Overall, these findings demonstrate that ESL/LM students rated production test formats (i.e., short answer/completion, restricted response, and extended essays) more favorably than recognition formats (i.e., true/false and multiple choice). In particular, ESL/LM students rated production items as being slightly more clear, interesting, straightforward, fair, valuable, and as holding higher expectancy of success than recognition formats. On the other hand, ESL/LM students also rated recognition formats as easier, more simple, less anxiety producing, and as producing more comfortable feelings than production formats on exams.

<table>
<thead>
<tr>
<th>Adjective Pairs</th>
<th>Production</th>
<th>Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>A. Difficult/Easy</td>
<td>3.50</td>
<td>1.34</td>
</tr>
<tr>
<td>B. Complicated/Simple</td>
<td>3.88</td>
<td>0.96</td>
</tr>
<tr>
<td>C. Unclear/Clear</td>
<td>4.58</td>
<td>1.05</td>
</tr>
<tr>
<td>D. Boring/Interesting</td>
<td>4.46</td>
<td>1.04</td>
</tr>
<tr>
<td>E. Tricky/Straightforward</td>
<td>4.66</td>
<td>1.10</td>
</tr>
<tr>
<td>F. Unfair/Fair</td>
<td>4.54</td>
<td>1.02</td>
</tr>
<tr>
<td>G. Worthless/Valuable</td>
<td>5.00</td>
<td>0.83</td>
</tr>
<tr>
<td>H. Low Expectancy of Success/</td>
<td>4.58</td>
<td>0.96</td>
</tr>
<tr>
<td>High Expectancy of Success</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I. High Anxiety Producing/</td>
<td>3.74</td>
<td>0.97</td>
</tr>
<tr>
<td>Low Anxiety Producing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J. Feeling Uncomfortable/Feeling</td>
<td>3.93</td>
<td>1.13</td>
</tr>
<tr>
<td>Comfortable with Exam</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.12: Mean Scores for Production and Recognition Formats by Adjective Pairs
Results exploring the relationship between preference for restricted-response formats (or multiple-choice formats) and actual performance on American Heritage restricted-response sections (or multiple-choice sections) of tests did not prove to be correlated (Research Question 10). In this regard, a student's attitude toward a particular test format does not appear to correlate with test performance using that particular format. In addition, analyses used to answer Research Question 11 showed that ESL/LM students' personal characteristics, such as length of residency, university experience, test knowledge, and previous courses, were not significantly correlated to students' preferences in testing.

Finally, regarding testing preferences, students were asked to indicate whether their preferences in testing differed in their L1 and L2. For true/false (60.5%), multiple choice (68.4%), short answer/completion (71.1%), and restricted response (65.8%), a clear majority of ESL/LM students felt their attitudes toward formats were the same in both languages. Students were more closely divided over extended essays; that is, only 55.3% of the students felt attitudes were the same in both languages.

Although 64% of the students, on average, felt their attitudes toward test formats was the same in both languages, 36% of these ESL/LM indicated that their attitudes differed by format. When students indicated a L1/L2 difference, they were asked to explain, or give a reason for the difference. A complete listing of the reasons offered by these ESL/LM students for the differences are provided in Appendix J; however, common responses will be highlighted for each format below. Overall, students most commonly identified vocabulary—wording, tricky words, or a limited vocabulary—as the factor that made their attitude toward a particular test format the same or different in their L1 and L2.

For the ESL/LM students (39.5%) who felt their attitudes toward true/false questions were different in their L1/L2, they explained that they considered true/false questions tricky, and easier to handle in their L1. For example, one student wrote that in his L1, "I would be able to understand the teacher's way of thinking. Sometimes those questions get
really tricky if I do not see things the way the teacher does," [01] Another student observed, "When American teachers try to be tricky with words, I get confused. No matter how good I am in English, I won't be as good as native speakers," [20].

Of the ESL/LM students (31.6%) who felt their attitudes toward multiple-choice questions differed in their L1/L2, they noted tricky wording as their main problem. For example, one student wrote, "Sometimes I find vocabulary I don't know in answers in English exams, and it makes it difficult to give right answers," [21]. Students also mentioned that multiple-choice questions were different for them in their L2 because of a focus on details, and use of much too similar, or far too many alternatives to choose among (e.g., A through I distractors).

Short-answer/completion questions were the format least likely to differ by language; however, the most frequent reason given for the difference attitudes was again vocabulary. One student described the problem as follows: "Many times you're asked for specific words. It may happen that you can't remember that one word in English, but you do remember it in your native tongue," [04]. Another student wrote, "I understand usually the concepts, but for this kind of question everything is centered in a single word that you can easily forget," [29].

ESL/LM students mentioned time limits, forgetting/memorizing, writing, and not being able to use dictionaries as some of the concerns they have with restricted-response formats in English. One student wrote, "In my language sometimes I can memorize and relate better because I have no limits in my expression," [05]. Another student said that without a dictionary, "I cannot express myself clearly enough. . . . I may know the answer but I just don't know how to say/write that," [18].

Of those students who felt extended-essay were different in English than in their native language, several reasons were given. Although students mentioned vocabulary limitations, they also were concerned with grammar, spelling, time to write, knowing how
to express themselves, and using an appropriate style. One student observed, "I always worry about grammar and vocabulary when I have an essay test. Sometimes I have hard time expressing what I really want to say," [06]. Another student felt that essays were simply easier in his first language: "Because in my language, I could feel free to express even with more words," [33].

Theoretical Model.

The purpose of this section of results is to present findings related to the theoretical model of potentially important variables associated with ESL/LM students' test performance in content-area classes (see Figure 3.1). In order to answer Research Question 12, and assess the influence of both exogenous and endogenous test-taker variables on test performance, a path analysis based on a multiple regression approach was carried out. The path analysis was conducted in three stages: 1) test performance was regressed on testing preferences, test-anxiety (i.e., defined as trait-anxiety for path analysis purposes), language group, and gender; 2) language proficiency was regressed on testing preferences, trait-anxiety, language group, and gender; and finally 3) test performance was regressed on all of the five independent variables.

Because this is a non-experimental design, and is descriptive in aim, path regression coefficients (i.e., unstandardized coefficients) were evaluated using two criteria of significance. First, a strong correlation is defined when a path regression coefficient is significant at \( p < .05 \). Second, a weak, but potentially important, correlation is defined when a path regression coefficient is significant at \( p < .10 \) but \( < .05 \). The relative strength of path regression coefficients is indicated by using a solid line for strong correlations \((p<.05)\), and a broken line for weak correlations \((p<.10 \text{ but } < .05)\). No line is provided when a correlation is below the \( p < .10 \) criterion. In path analysis diagrams, however, path coefficients (i.e., standardized or Beta weights), which are expressed as z-score, are used...
to clarify the importance of each variable in comparison to another. Results are presented by stage, and culminate in the full model test of significance.

**Stage One.** Results in this stage of the path analysis are presented as three multiple regressions: 1) regression of trait-anxiety on language group and gender; 2) regression of testing preferences (i.e., both overall scores and production/recognition scores) on language group and gender; and finally 3) regression of test performance on testing preferences, trait-anxiety, language group, and gender.

Table 5.13 presents the unstandardized and standardized (Beta) regression coefficients, and their accompanying t-values for the regression related to trait-anxiety. Results show that language group and gender together account for 17% of the variance in students' trait-anxiety.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>B</th>
<th>Beta</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Group</td>
<td>7.13</td>
<td>0.39</td>
<td>2.50*</td>
</tr>
<tr>
<td>Gender</td>
<td>1.08</td>
<td>0.06</td>
<td>0.38</td>
</tr>
</tbody>
</table>

\[ R^2 = .17, \ F(2,34) = 3.38* \]

*p < .05

Table 5.13: Multiple Regression Analysis of Trait-Anxiety on Language Group, and Gender

Results also show trait-anxiety among students did not differ significantly by gender; however, ESL/LM students' trait-anxiety differed significantly by language groups, \( r(36) = .40, p < .01, r^2 = .16 \). Romance language students (\( M = 65.40, SD = 8.33 \)) had higher levels of trait-anxiety than their non-romance language peers (\( M = 58.12, SD = 8.64 \)). Figure 5.1 presents the path coefficients (i.e., Beta weights) representing the regression of trait-anxiety on language group and gender, showing that only the path from
language group to trait-anxiety is substantiated. Eighty-three percent of the variance in trait-anxiety, however, remains unexplained by this model.

Table 5.14 presents the relevant statistics for the regression of testing preferences on language group and gender. Together language group and gender account for 17% of the variance in students' testing preferences, and both proved to be significantly associated to testing preferences.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>B</th>
<th>Beta</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Group</td>
<td>-2.94</td>
<td>-0.28</td>
<td>-1.75*</td>
</tr>
<tr>
<td>Gender</td>
<td>3.30</td>
<td>0.32</td>
<td>2.00**</td>
</tr>
</tbody>
</table>

$R^2 = .17, \ E(2.33) = 3.33**$

*$_{p < .10}, **_{p < .05}$

Table 5.14: Multiple Regression Analysis of Testing Preferences on Language Group and Gender
Using the p < .10 criterion, language group proved to be a weak path regression coefficient in association to students' testing preferences in the present study, $r (35) = -.26$, $p = .12$, $r^2 = .07$. Non-romance language group students ($M = 43.73$, $SD = 4.32$) were more positive toward all testing formats on average than romance language group students ($M = 40.99$, $SD = 5.69$). In analyses of each of the five test formats, only short-answer/completion questions differed by language group, $r (37) = -.29$, $p < .05$, $r^2 = .08$. Short-answer/completion questions were viewed more favorably by non-romance language students ($M = 48.21$, $SD = 8.17$) than romance language students ($M = 42.54$, $SD = 9.13$).

As the Beta coefficients indicate, gender proved to be more important than language group in understanding testing preferences, $r (35) = .30$, $p = .07$, $r^2 = .09$. Gender met the criterion for being a strong path regression coefficient in this analysis. Females ($M = 43.70$, $SD = 5.39$) had a more favorable attitude toward all testing formats on average than males ($M = 40.56$, $SD = 4.80$). Analyses of each of the five test formats also revealed that short-answer/completion formats were viewed much more favorably by females ($M = 49.02$, $SD = 7.35$) than males ($M = 41.53$, $SD = 9.14$), $r (37) = -.29$, $p < .05$, $r^2 = .08$.

Testing preferences were also analyzed by language group and gender using composite scores for production and recognition test formats. Students' preferences for recognition test formats did not correlate significantly with language groups or gender; however, preference for production test formats did. Table 5.15 presents the relevant regression statistics. Together language group and gender account for 20% of the variance in preference for production formats.

Females ($M = 45.37$, $SD = 7.06$) viewed test formats that required them to write/produce answers more favorably than males ($M = 40.95$, $SD = 7.30$), $r (37) = .30$, $p < .05$, $r^2 = .09$. ESL/LM students in the non-romance language group ($M = 45.41$, $SD$
Table 5.15: Multiple Regression Analysis of Production Format Preferences on Language Group, and Gender

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>B</th>
<th>Beta</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Group</td>
<td>-4.83</td>
<td>-0.33</td>
<td>-2.15*</td>
</tr>
<tr>
<td>Gender</td>
<td>4.95</td>
<td>0.34</td>
<td>2.21*</td>
</tr>
</tbody>
</table>

\[ R^2 = .20, \ F (2, 35) = 4.28* \]

* = \( p < .05 \)

Figure 5.2 presents the path coefficients (i.e., \( \beta \) coefficients) connecting gender to testing preferences (criterion \( p < .05 \)), and language group to testing preferences (criterion \( p < .10 \)). For this model, 83% of the variance in testing preferences remains unexplained by gender and language group.

**Figure 5.2:** Stage One path diagram: Influence of language group and gender on testing preferences.
The full model for Stage One required the regression of test performance on testing preferences, trait-anxiety, language group, and gender. Table 5.16 presents the relevant regression statistics. Results show that together preferences, trait-anxiety, group and gender account for 57% of the variance in ESL/LM students' test performance. Results also show that trait-anxiety and gender have a significant influence on test performance.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>B</th>
<th>Beta</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Testing Preferences</td>
<td>-.07</td>
<td>-.04</td>
<td>-0.29</td>
</tr>
<tr>
<td>Trait-Anxiety</td>
<td>-.41</td>
<td>-.57</td>
<td>-4.02**</td>
</tr>
<tr>
<td>Language Group</td>
<td>-3.09</td>
<td>-.23</td>
<td>-1.62</td>
</tr>
<tr>
<td>Gender</td>
<td>5.67</td>
<td>.42</td>
<td>3.04*</td>
</tr>
</tbody>
</table>

R² = .56,  F (2, 35 ) = 9.00**

* p < .005.  ** p < .001

Table 5.16: Multiple Regression Analysis of Test Performance on Testing Preferences, Trait-Anxiety, Language Group, and Gender

As indicated by beta coefficients, trait-anxiety is the most important variable in accounting for variance in test performance, r (36) = -.60, p < .001, r² = .36. Gender is the second most important variable although the zero-order correlation between gender and test performance is not statistically significant. On average, females (M = 83.78, SD = 5.54) scored higher on American Heritage tests than males (M = 80.12, SD = 7.40).

In the American Heritage context, the data suggest the paths connecting gender and trait-anxiety to content-area test performance are viable. However, testing preferences and language group did not produce significant path coefficients. Figure 5.3 presents the paths among independent and dependent variables which achieved statistical significance in Stage One of the path analysis.
Figure 5.3: Stage One path diagram: Influence of testing preferences, trait-anxiety, language group and gender on test performance.
Stage Two. For this stage of the path analysis, language proficiency was regressed on testing preferences, trait-anxiety, language group, and gender. Results of the analysis were not significant. Testing preferences, trait-anxiety, language group, and gender did not significantly account for any variance in students' language proficiency test performance. Paths connecting trait-anxiety and testing preferences to language proficiency are unsubstantiated by the data. Figure 5.4 presents the path diagram to this point of the analysis.

\[ p_{41} = 0.32 \]
\[ p_{32} = 0.39 \]
\[ p_{42} = -0.28 \]

Figure 5.4: Stage Two path diagram: Influence of testing preferences, trait-anxiety, language group, and gender on language proficiency.
Stage Three. In this stage of the path analysis, content-area test performance (i.e.,
a composite of the four American Heritage exams) was regressed on all of the independent
variables: language proficiency, testing preferences, trait-anxiety, language group, and
gender. Table 5.17 presents the means, standard deviations, and correlation matrix related
to these variables. Results indicate that none of the independent variables are highly
correlated to each other, and tests of multicollinearity were not significant.

<table>
<thead>
<tr>
<th>Variables</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Proficiency</td>
<td>105.92</td>
<td>3.74</td>
<td>6.85</td>
<td>-.13</td>
<td>.93</td>
<td>6.79</td>
</tr>
<tr>
<td>Testing Preferences</td>
<td>.07</td>
<td>33.98</td>
<td>-20.45</td>
<td>-.92</td>
<td>.56</td>
<td>12.62</td>
</tr>
<tr>
<td>Trait-Anxiety</td>
<td>.07</td>
<td>-.36</td>
<td>93.93</td>
<td>1.54</td>
<td>.80</td>
<td>-43.71</td>
</tr>
<tr>
<td>Language Group</td>
<td>-.03</td>
<td>-.31</td>
<td>.32</td>
<td>.25</td>
<td>.03</td>
<td>-1.29</td>
</tr>
<tr>
<td>Gender</td>
<td>.17</td>
<td>.19</td>
<td>.16</td>
<td>.11</td>
<td>.26</td>
<td>.67</td>
</tr>
<tr>
<td>Test Performance</td>
<td>.09</td>
<td>.29</td>
<td>-.62</td>
<td>-.35</td>
<td>.18</td>
<td>52.38</td>
</tr>
<tr>
<td>M</td>
<td>77.79</td>
<td>42.56</td>
<td>61.47</td>
<td>-----</td>
<td>-----</td>
<td>81.38</td>
</tr>
<tr>
<td>SD</td>
<td>10.29</td>
<td>5.83</td>
<td>9.69</td>
<td>-----</td>
<td>-----</td>
<td>7.24</td>
</tr>
</tbody>
</table>

Table 5.17: Means, Standard Deviations, Correlations, Variances, and Covariances for
Language Proficiency, Testing Preferences, Trait-Anxiety, Language Group, Gender, and
Testing performance

Table 5.18 presents the results of the regression analysis testing the influence of the
independent variables on test performance (i.e., the full model). Overall, the independent
variables accounts for 60% of the variance in content-area test performance. As found in
Stage One of this path analysis, the beta coefficients show that trait-anxiety is the most
important variable in accounting for variance in test performance, and gender is the second
Table 5.18: Multiple Regression Analysis of Content-Area Test Performance on Language Proficiency, Testing Preferences, Trait-Anxiety, Language Group, and Gender

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>B</th>
<th>Beta</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Proficiency</td>
<td>0.09</td>
<td>0.14</td>
<td>1.08</td>
</tr>
<tr>
<td>Testing Preferences</td>
<td>0.11</td>
<td>0.07</td>
<td>0.46</td>
</tr>
<tr>
<td>Trait-Anxiety</td>
<td>-0.41</td>
<td>-0.57</td>
<td>-4.03**</td>
</tr>
<tr>
<td>Language Group</td>
<td>-1.91</td>
<td>-0.14</td>
<td>-0.93</td>
</tr>
<tr>
<td>Gender</td>
<td>5.13</td>
<td>0.38</td>
<td>2.72*</td>
</tr>
</tbody>
</table>

\[ R^2 = .60, \ F (5, 26) = 7.73** \]

*p < .01  **p < .001

Table 5.18: Multiple Regression Analysis of Content-Area Test Performance on Language Proficiency, Testing Preferences, Trait-Anxiety, Language Group, and Gender

most important variable. Language proficiency, testing preferences, and language group did not produce significant path coefficients.

In the American Heritage context, the data suggest that only the paths connecting gender and trait-anxiety to content-area test performance are viable. Figure 5.5 presents the theoretical model with only those paths among independent and dependent variables which achieved statistical significance.

In order to determine the unique contribution of each set of variables in the three stages of the path analysis, the change in \( R^2 \) was calculated, with the accompanying change in the \( F \) statistic. Table 5.19 presents the change in \( R^2 \) at each stage of the path analysis. Results show that language proficiency only explains 3% of the variance in test performance, and is not significant. On the other hand, testing preferences and trait-anxiety together account for a statistically significant 45% of the variance in test performance when added to the model already containing language proficiency. Finally, language group and gender
Figure 5.5: Stage Three path diagram: Influence of language proficiency, testing preferences, trait-anxiety, language group, and gender on content-area test performance in American Heritage 100.
explain 15% of the variance in test performance—not statistically significant—when added to the model containing language proficiency, testing preferences, and trait-anxiety.

<table>
<thead>
<tr>
<th>Steps</th>
<th>$R^2$</th>
<th>Change in $R^2$</th>
<th>F-Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Proficiency</td>
<td>.03</td>
<td>.03</td>
<td>1.16</td>
</tr>
<tr>
<td>2 Proficiency, Preferences, Anxiety</td>
<td>.48</td>
<td>.45</td>
<td>3.74*</td>
</tr>
<tr>
<td>3 Proficiency, Preferences, Anxiety, Group, Gender</td>
<td>.60</td>
<td>.15</td>
<td>1.26</td>
</tr>
</tbody>
</table>

*$_p < .05$*

Table 5.19: Multiple Regression Analysis of Full Model Showing Change in $R^2$
Discussion of Findings

The results of this course-specific investigation provide an empirical account of content-area testing from a second language perspective. The interview and quantitative findings emerging from this American Heritage course are summarized, integrated, and discussed in light of current SLA, test anxiety, and psychometric literature. Following this discussion, the implications, limitations, and future directions for content-area testing research are presented.

1. Language Proficiency. As demonstrated by the results of this study, ESL/LM students' English language proficiency is not significantly correlated to students' content-area test performance. Although much has been written about ESL/LM students' struggle to linguistically represent what they know on content-area tests (e.g., LaCelle-Peterson and Rivera, 1994; Rosenthal, 1996; Short, 1993), the linguistic jeopardy in content-area testing is not quantitatively substantiated for these university-level ESL/LM students. This finding is consistent with Educational Testing Service (as cited in Graham, 1987) prediction that students who have demonstrated enough English language ability to be admitted for university work should not expect their English abilities to be "a major factor in influencing their academic success," (p. 515).

Nevertheless, Graham reports on several studies that create an unclear relationship, with a wide range of correlations, between English language proficiency and academic success. Part of the lack of clarity has been blamed on the inexactness of GPA as a measure. Academic success and language proficiency measures in these other studies have been several steps removed from where the problem in content-area testing for ESL/LM students takes place; that is, in the content-area classroom.

In the present study, ESL/LM students' language proficiency and test performance were measured concurrently, and analysis demonstrates that the influence of language proficiency was still negligible in this context. It is essential to underscore the importance
of context for several reasons however. First, the range of these students' language proficiency is restricted to the upper limits of the proficiency scale (e.g., Graham, 1987; Light, Xu, Mossup, 1987). Second, a majority of these students were international, not immigrant, students; and in this classroom outnumbered their English-speaking classmates. Third, student feedback indicates that the professor's instructional and testing practices effectively eliminated any language problem in testing for these second language students (e.g., Short, 1994).

Perhaps, in classes where ESL/LM students are in the minority, forced to express their content-knowledge within a 50-minute time frame, or tested using only one test format (e.g., only extended-essays or only multiple-choice items), the linguistic problems in testing for ESL/LM students would be more apparent. Language-related issues in testing remain worthy of investigation.

The findings of this study also suggest that educators should not only be asking if ESL/LM students experience language obstacles in testing, but rather under what conditions language proficiency hinders students' performance on content-area tests. As intimated by Steward (1991), Soto (1992), and Good (1987), focusing on language limitations may be a handy excuse for educators who are unwilling to make appropriate accommodations in testing practices for the diverse learners they face in the classroom.

ESL/LM students' proficiency in the individual skills of reading, writing, grammar or vocabulary do not predict content-area test performance. Still, ESL/LM students descriptively indicate that reading, writing, and vocabulary language skills are important to their success in content-area testing. This is consistent with the qualitative findings presented in the literature on ESL/LM students' in content-area classes (e.g., Horowitz, 1986; Leki & Carson, 1994; Leki, 1995). However, Leki's and Carson's (1994) claim to the singular importance of vocabulary knowledge in academic achievement is not quantitatively supported in the present study.
In summary, language problems in content-area testing may materialize more dramatically with students with lower levels of language proficiency, in classes where ESL/LM students are in the minority, and in classes where the instructor fails to adjust teaching and testing practices to a linguistically diverse student population. The present study also validates that linguistic obstacles in content-area testing for ESL/LM students can be diminished, if not overcome, in some contexts.

2. Test Anxiety. The results of this study suggest that students' level of trait-anxiety has more influence on test performance than levels of state-anxiety. The results of this study are consistent with Phillips et al.'s (1980) and Sapp's (1993) findings that ESL/LM students have significantly higher levels of trait-anxiety than language-majority students. Indeed, ESL/LM students possessed large amounts of both debilitating and facilitative anxiety; these two constructs do not have to be inversely related (Alpert & Haber, 1960). Unlike predictions by Madsen and Murray (1984), or Anderson and Sauser (1995), results do not support a gender difference in levels of trait-anxiety; rather a language group difference was detected similar to findings by Scott (1986).

In the present study, romance language students were found to be more prone to test anxiety than non-romance language peers. Although the reasons for these differences are unclear in this study, Phillips et al.'s (1980) argument that Mexican children are less comfortable with testing systems that stress competition and individualism could apply in this context. In future studies, language group differences could be explored more fully.

Overall, ESL/LM students' trait-anxiety was negatively and significantly correlated to test performance. In addition, students' familiarity with the American testing system was inversely and moderately correlated to trait-anxiety. These findings imply that ESL/LM students could benefit from instruction aimed at lowering the emotional component of test anxiety (Sapp, 1993), and at increasing students' cognitive awareness of test-taking strategies; that is, test wisdom (Anderson & Sauser, 1995).
In the present study, a lack of consistent correlations between test performance and state-anxiety provides additional evidence that the professor's accommodations in testing (i.e., no time limits, allowing the use of dictionaries, etc.) helped to minimize the situation-specific, or task-dependent aspects of test anxiety for these students. This finding suggests that professors can create testing conditions so positive as to diminish students' state-anxiety; therefore, reducing the influence of anxiety on content-area test performance for ESL/LM students.

3. Testing Preferences. In the American Heritage context, ESL/LM students scored higher overall on restricted-response questions than multiple-choice questions. ESL/LM students felt the restricted-response format was the most difficult section of the test, and it generated the most state-anxiety in students. Nevertheless, students also felt that restricted-response formats best allowed them to demonstrate their content knowledge. This finding supports Nield's and Wintre's (1986) research with language-majority students, who also most strongly prefer restricted-response formats.

Many of the findings related to ESL/LM students' testing preferences parallels research with language-majority populations (Gellman & Berkowitz, 1993; Zeidner, 1987). For example, ESL/LM students' attitudes toward true/false, multiple-choice, short-answer/completion, restricted-response and extended-essay test formats differ significantly; however, the evidence in the present study requires that judgments about where those differences lie be suspended, earmarking this aspect of testing preferences as worthy of subsequent research and replication (Keppel, 1991).

Several other findings regarding students' general testing preference in content-area testing are noteworthy. First, ratings show that ESL/LM students are most positively disposed toward short-answer/completion formats, followed in order of preference by restricted-response, multiple-choice, true/false, and extended-essay formats. Results also show a gender and language group difference in students' preference for short-
answer/completion tests. Although more experimental research is needed, it appears that language-majority populations 1) do not rate short-answer/completion items as positively, or 2) dislike essay formats as much as ESL/LM students (e.g., Nield & Winter, 1986).

Second, ESL/LM students generally believe that they can demonstrate their contentknowledge on tests better with test formats that require answer production over answer recognition. Similar to findings among language-majority students, ESL/LM students consider recognition formats as easier, more simple, more comfortable, and the least anxiety-producing. Interview and quantitative data from the American Heritage setting lend further support to research by Foos (1992), Gellman and Berkowitz (1993), and Nield and Wintre (1986), which suggests that students believe it is easier to do better--score higher--on recognition formats.

The findings of this study are also consistent with research intimating gender differences in testing preferences (e.g., Anderson, 1989; Gellman & Berkowitz, 1993). Among these ESL/LM students, females were found to prefer production test formats more than males. The preference for short-answer/completion questions, however, is unique to these ESL/LM students. Perhaps, this preference for short-answer/completion items reflects how ESL/LM students are used to being tested in English; namely, this format is typical in many English language textbooks, as well as on teacher-constructed, and standardized language tests.

This study supports the claim qualitatively and quantitatively that testing practices in the American testing system may not be consonant with the assumptions and expectations of ESL/LM students' previous, home-country experience (Lam, 1993; Sarason, 1980; Scott, 1986). ESL/LM students were able to identify testing practices which differed from their native experience. ESL/LM students' reactions to multiple-choice test formats were also not nearly as negative as advocates for ESL/LM students depict (e.g., Anderson, 1988; Lam, 1993; Lemke, 1990; Soto, 1992). These university-level ESL/LM students.
considered multiple-choice formats easy, interesting, and the least anxiety-producing of all formats. The bias in standardized testing addressed in the literature needs to be clarified given these students' greater comfort with recognition formats. Perhaps, there is greater jeopardy to the ESL/LM student population in not being tested at all over being exposed to standardized, multiple-choice test formats.

Findings also suggest a slight language group difference in testing preferences, with non-romance students being more positive toward all test formats than romance language students. Further research is needed to delineate if this language group difference is based on cultural views of learning, race, international versus immigrant student status, or socioeconomic (LaCelle-Peterson & Rivera, 1994; Lemke, 1990; Soto, 1992).

Finally, these ESL/LM students indicated that their attitudes toward test formats were the same in their L1 and L2. Of those students who felt their attitudes differed across languages, the most common reason given by students for that difference was a limited vocabulary. In addition, students also listed time limits, memorizing, writing, and dictionary use as other reasons for viewing various test formats differently in their L1 and L2. These findings are consistent with the ESL literature suggesting that problems in testing for ESL/LM students arises from a lack of appropriate accommodation by teachers (e.g., Brinton et al., 1989; Latulippe & Light, 1983; Kinsella, 1992; Rosenthal, 1996).

In summary, these findings intimate that ESL/LM students' attitudes toward test formats vary significantly. ESL/LM students' preferences in testing tentatively appear to parallel findings reported for language-majority populations. ESL/LM students suggest that formats that require recognition of correct answers are easier, but that production formats are more fair, straightforward, and valuable. In the American Heritage course, ESL/LM students showed a strong preference for, and scored more points on, the restricted-response sections of their tests; however, when asked which format they preferred generally, one-third of the students selected the multiple-choice format. More research is
need to clarify when, and under what conditions, ESL/LM students prefer production over recognition test formats.

**General Implications**

The purpose of this study was to investigate content-area testing from a second language perspective. Although this study was exploratory, the quantitative findings gathered in the American Heritage context suggest several implications for the advancement of theory, pedagogy, and research aimed at the growing numbers of language-minority students entering American colleges and universities. These general implications follow.

First, the theoretical model formulated to account for content-area test performance—based on the constructs of language proficiency, test anxiety, and testing preferences—has been tentatively substantiated. To date, problems in testing for ESL/LM students have been discussed most frequently in terms of limited English proficiency (e.g., Anderson, 1988; Simms & Leonard, 1986). This language focus may be justified for students with low or intermediate levels of English language proficiency. For university-level ESL/LM students, however, conceptualization of the obstacles in testing should be extended to include facets of test anxiety and testing preferences. Taken together, the qualitative and quantitative results suggest that the disadvantage in content-area testing for ESL/LM students is more accurately understood as a multi-dimensional phenomenon, comprised of factors both within and outside the learner’s control.

Second, trait-anxiety proved to be the variable accounting for the most variance in ESL/LM students’ content-area test performance. It appears greater attention on what teaching and testing practice allay test anxiety among ESL/LM students is justified.

Third, becoming and remaining competitive with language-majority peers requires ESL/LM students to overcome language, anxiety, or format preference obstacles in content-area testing. Based on learner initiative or ESL program support, each of these obstacles are amenable to instruction. Nevertheless, this classroom-based study lends empirical
support to the idea that instructional and testing accommodations on the part of teachers can do much toward leveling the playing field for ESL/LM students in content-area courses.

Fourth, ESL/LM students prefer production test formats for best demonstrating their content knowledge although they view recognition formats as easier to do well on. Under what conditions students prefer production formats or recognition formats is not evident in the present study. More research is also needed to determine empirically whether ESL/LM students' preferences in testing significantly differ from language-majority students' preferences.

Fifth, in this American Heritage classroom, the professor effectively minimized state-anxiety, and any evidence of a second language disadvantage among his ESL/LM students. Continued efforts to collaborate with greater numbers of faculty across the disciplines on behalf of ESL/LM student populations appears warranted; however, ESL teacher education programs have an obligation to prepare their students for effective cross-disciplinary collaboration, which entails both learning from and sharing with other faculty.

General Limitations

In this classroom-based, non-experimental, and exploratory research design, causality was not at issue. The aim of the study was to illuminate the obstacles in testing for ESL/LM students in a naturally occurring and intact group, using both quantitative and qualitative data. Pivotal to the design was measuring language proficiency, test anxiety, and academic achievement simultaneously. Finding a large number of ESL/LM students in one university-level, content-area course is rare. The American Heritage setting at Brigham Young University, with 38 ESL/LM students, met the context requirements for this study. Randomization and treatment manipulation were not employed; therefore, the context also limits the interpretation that can be applied to these data.

First, the findings presented in this study explain content-area test performance in the American Heritage classroom. Although findings reveal the challenges ESL/LM students
experience in content-area classes, the findings cannot be generalized to other ESL/LM student populations, or disciplines without further research employing experimental or quasi-experimental designs. It is important to remember that the American Heritage professor taught and tested in ways that supported the needs of ESL/LM students. In other content-area courses, ESL/LM students may, indeed, experience greater language, anxiety, or format problems in expressing their knowledge on content-area tests.

Multiple regression and path analysis approaches are typically large sample statistical procedures, which lend support to causality in experimental designs. In the present study, multiple regression was selected because it does not require 1) equal sample sizes, 2) distortion of the data (i.e., dichotomizing continuous variables), or 2) orthogonal variables. For non-experimental designs, multiple regression is considered more powerful than analysis of variance alternatives (Keppel & Zedeck, 1989; Sarason, 1980). This study, however, only allows prediction within the sample. The statistical assumptions of multiple regression would need to be met to claim external validity.

The findings of this study are descriptive and correlational regarding content-area testing for ESL/LM students. Nevertheless, the use of triangulation of data sources and methods lends support to the findings put forth in this study. Further research is needed, using larger sample sizes and randomization, to improve generalizability to other ESL/LM populations, and content-areas.

Future Directions

The findings of this study tentatively demonstrate that given 1) advanced language skills, 2) high levels of test anxiety, 3) a variety in test formats, and 4) a student's gender and language background, a substantial amount of the variance in content-area test performance can be accounted for in the classroom setting. These findings suggest several different avenues of research would be worth pursuing.
Further research is needed to empirically document under what conditions language proficiency influences content-area test performance. As Vandrick (1995), Hernández (1992), and Graham (1987) suggest, the ESL/LM population is not homogeneous. Test anxiety and testing preferences should be investigated contrasting international and immigrant populations, and groups by languages (e.g., Spanish speakers and Chinese speakers).

Further survey research comparing native and non-native speaker preferences for various test formats is also reasonable, considering the continuing debate concerning the efficacy of standardized tests. In the university setting, comparisons among essay, restricted-response, and multiple-choice formats appear to be the most valid.

Conclusion

In order to explore what, if any, disadvantages ESL/LM students experience in content-area testing, a theoretical model was formulated based on the constructs of language proficiency, test anxiety, and testing preferences. In the present study, these variables were found to accounted for 60% of the variance in ESL/LM students' content-area test performance. The quantitative findings support the notion that there are gender differences in ESL/LM students' testing preferences. In addition, language group differences were found related to test anxiety and testing preferences. In this American Heritage context, no relationship between language proficiency and test performance was documented.

These findings add quantitative evidence to the understanding of content-area testing from a second language perspective. In Chapter 6 the qualitative and quantitative portions of this study are synthesized, highlighting points of corroboration and divergence among data sources.
CHAPTER 6

CONCLUSIONS

Introduction

The purpose of this study was to examine the role of language proficiency, test anxiety, and testing preferences in ESL/LM students' test performance in a content-area course. The qualitative interview results reported in Chapter 4 provided an overview of the general concerns, problems, and preferences ESL/LM students have in test situations in university content-area courses. The quantitative findings presented in Chapter 5 provided a course-specific account of ESL/LM student test performance in an American Heritage course. The purpose of this chapter is to synthesize the general and specific, and the qualitative and quantitative into a second language account of content-area testing for ESL/LM students. Following a synthesis of findings, relevant implications, limitations, and future directions will be addressed related to theory, practice, and research.

Synthesis of Findings

The research questions guiding this study allowed content-area testing to be explored from four perspectives: 1) problems and practices; 2) language proficiency; 3) test anxiety; and 4) testing preferences. The qualitative interview data in both general and course-specific settings provides a useful profile of testing problems and practices that concern ESL/LM students. This profile is presented first, contextualizing the subsequent synthesis of qualitative and quantitative findings related to language, anxiety, and preferences.
1. The Qualitative Profile: Problems and Practices. ESL/LM students described their experiences in content-area testing from two different perspectives: one general, and the other course-specific. Students in both contexts supplied direct and indirect evidence of problems in testing related to context, expectations, and language. Students also intimated which in-class and test-preparation practices of their teachers were helpful or harmful to them as ESL/LM students.

In the broadest terms, the students who participated in the general interviews provided the framework against which the course-specific student feedback can be assessed. Table 6.1 presents a profile of the general and course-specific findings, allowing four generalizations to be made regarding the American Heritage setting.

First, the classroom context for many ESL/LM students is problematic. In general, ESL/LM students perceive their minority status, feel grades do not adequately reflect their knowledge, question grading practices of teachers, and are often put in the position of training their faculty in how to recognize their needs. In the American Heritage context, the ESL/LM students did not have to contend with these obstacles. They were in the majority, felt their grades reflected their knowledge, and felt the teacher was sensitive to their needs in teaching and grading. Overall, the American Heritage context was positive for ESL/LM students.

Second, ESL/LM students in both interview contexts acknowledged differences between their home-country and American testing experiences and expectations. These students concurred that the American testing system required more/longer tests, valued personal experience, and felt easier than their home-country systems. They also found take-home tests, testing centers, class presentations, and using drawings on tests as unique to their American experience. On these points there was agreement among interview subjects.
### Problems in Testing

<table>
<thead>
<tr>
<th>A. Context</th>
<th>General Interview</th>
<th>American Heritage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Minority in class</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>2. Grades accurately reflect</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>3. Grading is fair</td>
<td>?</td>
<td>Y</td>
</tr>
<tr>
<td>4. Teachers sensitive to needs</td>
<td>?</td>
<td>Y</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Testing differs from home-country</th>
<th>Y</th>
<th>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. More tests</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>2. Personal opinions count</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>3. Professors care more</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>4. System is easier</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>5. New alternatives in testing</td>
<td>Y</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C. Language-related problems</th>
<th>Y</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Vocabulary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Difficulty understanding questions</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>3. Difficulty memorizing</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>4. Knowing how to write</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>5. Time limitations</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>6. Trapped knowledge</td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>7. Keeping pace</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

### Helpful Teaching Practices

<table>
<thead>
<tr>
<th>A. In-class requests</th>
<th>N</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Don't speak fast</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Explain difficult concepts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Reinforce what is in book</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Discuss format of upcoming test</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Provide specific feedback on improving</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Have office hours</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>7. Allow extra credit</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>8. Have smaller classes</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Test-preparation requests</th>
<th>N</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Write test questions clearly</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Use simple vocabulary</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Allow more time during tests</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Allow dictionary use</td>
<td>?</td>
<td></td>
</tr>
<tr>
<td>5. Have someone to help explain difficult test questions available</td>
<td>N</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Y = Yes; N = No; ? = Sometimes.

Table 6.1: General Interview and American Heritage Profiles of Content-area Testing
Third, students participating in the general interviews described a broad range of language-related problems in testing as listed in Table 6.1. American Heritage students did not. Only one American Heritage student noted problems memorizing large amounts of content; she did not intimate that the difficulty was memorizing in English. These American Heritage students did, however, describe language-related problems in other content-area classes. Overall, language-related problems in testing were substantiated, but apparently minimized in the American Heritage setting.

Fourth, general interview subjects described several helpful in-class and test-preparation practices which their content-area teachers engaged in with a lack of consistency. The American Heritage students, however, described their professor as sensitive to their second language needs: using clear language, supporting lectures by writing on the blackboard, providing practice tests, allowing dictionary use, and not using time limits. American Heritage students did note a need for more help in making sense of their reading assignments from the textbook. Overall, it appears that the American Heritage professor accommodated his ESL/LM students by using helpful instructional and testing practices.

These two sources of qualitative feedback substantiate one another regarding the problems and helpful practices related to content-area testing. These data were gathered to complement, and enhance interpretation of the quantitative outcomes. To this end, it is important to observe that the American Heritage context, where the quantitative data were gathered, appeared to be ideal in reducing the obstacles for ESL/LM students in demonstrating their content knowledge.

2. Language Proficiency. In the American Heritage context, ESL/LM students' language proficiency did not empirically account for test performance outcomes, suggesting that no language barrier existed in students' content-area testing experience. This no-barrier outcome could suggest that language problems for ESL/LM students in university
content-area courses has been exaggerated; however, students' qualitative feedback tempers this interpretation. ESL/LM students were not only able to describe the language-related problems they have in testing, they were also able to articulate the teacher practices that minimize those problems (e.g., dictionary use). The no-barrier outcome for language proficiency in the American Heritage setting can be attributed—to some degree—to the professor's instructional and testing accommodations. This study provides empirical evidence of how content-area testing can be made more equitable for ESL/LM students.

3. Test Anxiety. The qualitative and quantitative data provide evidence of the importance of test anxiety as a variable influencing testing outcomes for ESL/LM students. In addition, this sample of ESL/LM students proved to possess higher levels of trait-anxiety than expected, compared to language-majority students. Students' state-anxiety during tests was relatively low, showing little association to test performance. Students' familiarity with the American system of testing was correlated to test anxiety; other personal characteristics were not predictive.

The qualitative data provide insights into these test anxiety findings. Students describe problems in testing related to language, culture, expectations, test conditions, and teacher practices. Any one of these concerns, or combination of concerns, could potentially explain heightened levels of test anxiety among ESL/LM students.

4. Testing Preferences. Findings related to ESL/LM students' preferences for true/false, multiple-choice, short-answer/completion, restricted-response, and extended-essay formats were not qualitatively and quantitatively consistent. When selecting among formats, a majority of ESL/LM students preferred multiple-choice tests. When asked to rate formats on the basis of bipolar adjectives, ESL/LM students were most positive about short-answer/completion, and restricted-response formats.
In the American Heritage context, students felt the restricted-response sections of tests allowed them to demonstrate their content knowledge best. Indeed, test performance data empirically confirmed that students scored higher on restricted-response formats than multiple-choice formats.

Quantitatively, neither overall format or specific format ratings proved to be significant in predicting students' test performance. Although students do rate these test formats differently, judgments about where those differences lie is unclear. Nevertheless, gender and language group did reveal patterns in testing preferences.

In the American Heritage context, teacher adjustments (e.g., no time limits, etc.) may have diminished the challenge various test formats pose in testing outcomes for ESL/LM students. Further research is needed to determine if under less accommodating circumstances testing preferences would predict testing outcomes for ESL/LM students.

Synthesis of Implications

These qualitative and quantitative findings extend and suggest new ways of conceptualizing, ameliorating, and investigating content-area testing for ESL/LM students. Although implications have been addressed in previous chapters, the aim of this synthesis is to underscore how these findings may alter and improve our understanding of the ESL/LM student experience in content-area testing.

1. Theoretical Implications. Theories of language proficiency, test anxiety, and testing preferences merge in examining the obstacles ESL/LM students face in expressing content knowledge in a second language. The qualitative and quantitative findings give cautious support to this conceptualization of testing for ESL/LM students. The findings suggest that problems in testing is a multidimensional phenomenon, with aspects internal and external to the student's control.

Internal factors reflects linguistic, cognitive, and affective concerns in content-area testing. These internal factors require ESL/LM students to improve their strategic
competence by developing linguistic, cognitive, and test-taking strategies. How effectively ESL/LM students develop these coping skills when the external factors have not been adjusted to their needs is a question of opportunity to learn, time to practice, and effort.

External factors rest in the control of content-area faculty, and relate to decisions made regarding pedagogical and testing practices; namely, vocabulary use, test format, time limitations, and grading practices. Adjustments to the external conditions of testing require sensitivity and effort on the part of content-area faculty. On the surface, the external factors appear easier to overcome in comparison to the internal dimension for ESL/LM students.

Overall, the language-anxiety-preference model of content-area testing was not fully substantiated by the data. More research is required to determine if language proficiency and testing preferences do influence test performance under less accommodating content-area conditions. The American Heritage professor provided empirical evidence of the efficacy of testing accommodations for ESL/LM students. Indeed, if the obstacles in testing for ESL/LM students are mostly an issue of adjusting external factors, faculty have an obligation to make content-area testing equal and equitable for ESL/LM students (e.g., Grant, 1992; McKay & Freedman, 1990).

2. Classroom Implications. Findings have clear implications for the content of ESL programs, the breadth of ESL teacher preparation programs, and the nature of collaboration with content-area faculty. These implications follow.

A. ESL Programs. Four implications can be drawn from the findings of this study for ESL programs charged with preparing students for the university classroom. First, intensive ESL programs should continue to develop curricula which distinguishes between academic and non-academic track courses. The expectations and goals of English for Academic Purposes (EAP) courses differ in kind from courses intended for social/interpersonal language skill development; however, the content of intensive program courses do not always reflect these goals.
Second, many ESL programs maintain a skill-based approach to EAP skills, offering academic reading, writing, or listening/speaking courses to matriculated ESL/LM students. Although useful, these courses may not go far enough in acclimating ESL/LM students to the demands of university testing. EAP programs should appropriately integrate content-area testing concerns and strategies into skills-based courses.

Third, because test anxiety is both an emotional and cognitive construct, it is amenable to instruction. Although trait-anxiety is stable, state-anxiety is task-dependent. Developing both general and format-specific test-taking strategies should be a priority, not a elective, for university-bound ESL/LM students.

Fourth, when students describe a linguistic problems in demonstrating knowledge in testing, they commonly draw attention to their limited vocabulary. If vocabulary ability were shown to be empirically important in test performance, would students benefit from more instruction geared toward academic vocabulary development? ESL programs are well-suited to address automaticity in academic vocabulary use in instruction.

B. Teacher preparation programs. The problems and concerns of ESL/LM students in content-area classes imply an agenda for ESL educators. Future ESL teachers must be prepared to teach both general and EAP courses. Although EAP literature is abundant, and the literature focusing on ESL/LM students in content-area classes is expanding, most teacher preparation programs do not prepare teachers to bridge the gap between the ESL and content-area classrooms. Teacher trainers must ask themselves if, and how, they are preparing ESL teachers to collaborate with, and be informed resources to content-area faculty.

C. Content-area Collaboration. These findings demonstrate the benefit to ESL/LM students when content-area faculty make appropriate adjustments in testing practices. ESL teachers and researchers should increase their efforts to publish in content-area journals, engage in more cross-disciplinary research, and make connections with content-area faculty.
in pre-service and in-service programs, on informal and formal basis. Such efforts would impact ESL programs, teacher preparation courses, and awareness among content-area faculty about the needs of ESL/LM students.

3. Research Implications. The present study has contributed to understanding content-area testing by providing baseline qualitative and quantitative evidence predicting test performance for ESL/LM students. The use of multiple methods and data sources in a classroom setting has improved data interpretation, and provided correlational insight into the role of language proficiency, test anxiety, and testing preferences in ESL/LM students' test performance. Students' qualitative feedback provided evidence of language-related problems in testing. The empirical evidence demonstrated how appropriate instructional and testing adjustments can diminish the effects of language proficiency and state-anxiety on testing outcomes for ESL/LM students. The value of qualitative and quantitative approaches to data collection and interpretation has been validated in the present study, and should be a consideration in future research.

Limitations

The limitations of this investigation have been outlined previously in Chapters 4 and 5. Although the descriptive and empirical evidence showed trends toward convergence, the internal and external validity of these data could be enhanced by further replication, larger sample sizes, randomization, and quasi-experimental designs.

Future Directions

Three directions in research are intimated by the outcomes of this study. Further insights into content-area testing can be gained by conducting large scale research comparing the testing preferences of language-majority and language-minority student populations. More investigation is warranted to determine under what testing conditions language proficiency, in particular vocabulary knowledge, hamper test performance. In addition, research and instructional collaboration with content-area faculty related to
learning and testing-taking tasks within content-area domains remains a potentially rich avenue for further research as well.

Conclusion

With the changing demographics of the United States, and a push to reform education through testing, ESL professionals have a vested interest in more fully understanding content-area testing from a second language perspective. As every teacher in every school district and university across the country begins to interact with greater numbers of ESL/LM students, the role of ESL professionals in supporting and advocating equality of outcomes for linguistically and culturally diverse learners will become more important. Arming our profession with the knowledge and ability to interface with content-area faculty across the disciplines should be a top priority in the twenty-first century. Against this backdrop, understanding how ESL/LM students come to accurately and consistently represent their content knowledge on tests will remain a second language issue worthy of systematic investigation.
APPENDIX A
THE RELATIONAL LEARNING PROCESS

152
THE RELATIONAL LEARNING PROCESS
APPENDIX B

SEMI-STRUCTURED INTERVIEW QUESTIONS

Background Information
1. What country are you from?
2. How long have you been in the United States?
3. How long have you studied at an American university?
4. What are you hoping to major in at Brigham Young University?

As I mentioned on the telephone, I am interested in your personal experiences at the university, particularly how you have experienced testing. I am trying to learn more about how ESL students, like yourself, adjust to the expectations of our testing system.

Primary Testing Concerns
5. Could you describe your first experiences in taking tests at an American university?
6. How do you think the types of tests you are given in the U.S. differ from the types of tests you would expect to take in your home country?
7. From your own experience, could you describe the kinds of problems you encounter when taking tests in English?
   a. What have you learned about taking tests in university courses?
   b. What kinds of strategies do you use when taking an exam?
   c. What kinds of strategies do you use in taking exams now, that you didn't use when you first arrived?
8. How well do you think the grades you get on tests reflect your abilities?
9. What kinds of tests do you prefer taking?
   a. Which test format do you consider the easiest for you to do well on?
   b. Which test format do you consider the most difficult for you?
   c. Do you believe your test preferences are the same in your first and second language?
10. Which language skill do you think is the most important in helping you best show your knowledge on tests?
Secondary Testing Concerns

11. In your opinion, how can professors help you to perform better on tests you take in their classes?

12. If you participated in an ESL program, how do you think your language teachers could help you prepare better for test-taking in university courses?
APPENDIX C
EXAMPLES OF LANGUAGE PROFICIENCY TEST ITEMS

Michigan Test of English Language Proficiency (MTELP)

Sample items taken from Form E:

Grammar:  "When did it begin snowing?"
          "It started ______ the night."
          a) during  
          b) by      
          c) from    
          d) at

          "Mr. Anderson is waiting to use this machine."
          "I suppose he thinks __________ early today."
          a) I were to leave
          b) that I to leave
          c) that I am leaving
          d) me to leave

Vocabulary: He turned and gave a bitter retort.
          a) reply
          b) note
          c) look
          d) appeal

          I couldn't hear her because she spoke so __________ .
          a) commonly
          b) barely
          c) smoothly
          d) faintly
Reading: The difference between biological and physical science is not that one is inexact, the other exact, but in degree of exactness, this being related to the number of variables which must be dealt with simultaneously and the extent to which they can be controlled. In general, the biological sciences must deal with larger errors than the physical sciences; but it is not uniformly true, as the student will realize when he considers the accuracy of meteorological prediction or if he comprehends the meaning of the fact that the structural engineer considers it necessary very often to use a safety factor of two or three percent. The statistical principles of dealing with error of measurement, or in prediction and generalization, are the same whether the errors are large or small. Statistics is not a means of confusing issues that would otherwise be clear, nor a substitute for obtaining clear answers, but a means of checking and controlling conclusions by providing an estimate of the error to which a conclusion is subject.

What do statistics do for science?
   a) They get rid of intuition.
   b) They make it more logical.
   c) They reduce everything to numbers.
   d) They given an idea of the among of error involved.

When the error is small, the statistical principles used in dealing with it are...
   a) simpler.
   b) more difficult.
   c) not yet understood.
   d) the same.
APPENDIX D

EXAMPLE WRITING PROMPTS AND RATING CRITERIA

Impromptu Writing Assessment

Directions: You have 30 minutes to write. You must write on only ONE of the two topics provided. Before you begin writing, you might want to take a few minutes to plan your composition, or make a short outline to organize your thoughts. After writing, you may want to save a few minutes to make changes or corrections to your composition.

Sample item:

1. What do you think is your country's greatest problem? Explain in detail, and tell what you think can be done about it.

2. What do you believe are the characteristics of good government? Explain and give examples.

3. Compare American colleges to colleges in your home country. How are they the same? How are they different? Where would you prefer to attend college, and why?

The MELAB Rating Levels:

97 Topic is richly and fully developed. Flexible use of a wide range of syntactic (sentence level) structures, accurate morphological (word forms) control. Organization is appropriate and effective, and there is excellent control of connection. There is a wide range of appropriately used vocabulary. Spelling and punctuation appears error free.

93 Topic is fully and complexly developed. Flexible use of a wide range of syntactic structures. Morphological control is nearly always accurate. Organization is well controlled and appropriate to the material, and the writing is well connected. Vocabulary is broad and appropriately used. Spelling and punctuation errors are not distracting.

87 Topic is well developed, with acknowledgment of its complexity. Varied syntactic structures are used with some flexibility, and there is good morphological control. Organization is controlled and generally appropriate to the material, and there are few problems with connection. Vocabulary is broad and usually used appropriately. Spelling and punctuation errors are not distracting.
83 Topic is generally clearly and completely developed, with at least some acknowledgment of its complexity. Both simple and complex syntactic structures are generally adequately used; there is adequate morphological control. Organization is controlled and shows some appropriacy to the material, and connection is usually adequate. Vocabulary use shows some flexibility, and is usually appropriate. Spelling and punctuation errors are sometimes distracting.

77 Topic is developed clearly but not completely and without acknowledging its complexity. Both simple and complex syntactic structures are present; in some “77” essays these are cautiously and accurately used while in others there is more fluency and less accuracy. Morphological control is inconsistent. Organization is generally controlled, while connection is sometimes absent or unsuccessful. Vocabulary is adequate, but may sometimes be inappropriately used. Spelling and punctuation errors are sometimes distracting.

73 Topic development is present, although limited by incompleteness, lack of clarity, or lack of focus. The topic may be treated as though it has only one dimension, or only one point of view is possible. In some “73” essays both simple and complex syntactic structures are present, but with many errors; others have accurate syntax but are very restricted in the range of language attempted. Morphological control is inconsistent. Organization is partially controlled, while connection is often absent or unsuccessful. Vocabulary is sometimes inadequate, and sometimes inappropriately used. Spelling and punctuation errors are sometimes distracting.

67 Topic development is present but restricted, and often incomplete or unclear. Simple syntactic structures dominate, with many errors; complex syntactic structures, if present are not controlled. Lacks morphological control. Organization, when apparent, is poorly controlled, and little or no connection is apparent. Narrow and simple vocabulary usually approximates meaning but is often inappropriately used. Spelling and punctuation errors are often distracting.

63 Contains little sign of topic development. Simple syntactic structures are present, but with many errors; lacks morphological control. There is little or no organization, and no connection apparent. Narrow and simple vocabulary inhibits communication, and spelling and punctuation errors often cause serious interference.

57 Often extremely short; contains only fragmentary communication about the topic. There is little syntactic or morphological control, and no organization or connection are apparent. Vocabulary is highly restricted and inaccurately used. Spelling is often indecipherable and punctuation is missing or appears random.

53 Extremely short, usually about 40 words or less; communicates nothing, and is often copied directly from the prompt. There is little sign of syntactic or morphological control, and no apparent organization or connection. Vocabulary is extremely restricted and repetitively used. Spelling is often indecipherable and punctuation is missing or appears random.
APPENDIX E

STATE-ANXIETY MEASURES
Test One Evaluation

NAME: ________________________ Section Number: ____________

Although I need your name for data analysis purposes, YOUR ANSWERS WILL BE KEPT CONFIDENTIAL!

Instructions:
Please tell me how you feel about each part of the exam you just completed. Put an X in one of the 5 spaces to show your agreement or disagreement with each statement.

**Part I: Write and Draw**

1. I liked this part.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

2. This part was difficult.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

3. I felt pleasant (happy, calm) during this part.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

**Part II: Multiple-Choice**

1. I liked this part.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

2. This part was difficult.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

3. I felt pleasant (happy, calm) during this part.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

**Part III: Pattern Model**

1. I liked this part.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

2. This part was difficult.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

3. I felt pleasant (happy, calm) during this part.  
   - Strongly Agree  
   - Agree  
   - Undecided  
   - Disagree  
   - Strongly Disagree

**Part IV: Rate your preparation and performance**

1. How knowledgeable do you consider yourself to be about taking tests at the university level in the U.S. system? (Put an X on the scale below.)
   - Not Very knowledgeable  
   - 1  
   - 2  
   - 3  
   - 4  
   - 5  
   - 6  
   - 7 Very knowledgeable

2. Circle the grade you believe you will receive ON THIS EXAM: (your best guess)
   - A  
   - A-  
   - B+  
   - B  
   - C+  
   - C  
   - C-  
   - D+  
   - D  
   - D-  
   - F

3. Which test format do you believe allowed you to best show what you know on this test? (Check ONLY ONE.)
   - Write & Draw  
   - Multiple choice  
   - Pattern Model
4. How effective do you think were in preparing for this test? (Put an X on the scale below.)

Not effective preparation 1-------2-------3-------4-------5-------6-------7 Very effective preparation

5. Briefly describe how you prepared for this exam.

6. Would you prepare any differently now that you have seen the type of exam your teacher gives?
**Test Two Evaluation**

**NAME:** ________________________  **Section Number:** ____________

Although I need your name for data analysis purposes, YOUR ANSWERS WILL BE KEPT CONFIDENTIAL!

**Instructions:**
Please tell me how you feel about each part of the exam you just completed. Put an X in one of the 5 spaces to show your agreement or disagreement with each statement.

<table>
<thead>
<tr>
<th>Part I: Write and Draw</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. This part was difficult.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part II: Multiple-Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
</tr>
<tr>
<td>2. This part was difficult.</td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part III: Pattern Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
</tr>
<tr>
<td>2. This part was difficult.</td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Part IV: Rate your preparation and performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Circle the grade you believe you will receive ON THIS EXAM: (your best guess)</td>
</tr>
<tr>
<td>A  A-  B+  B  B-  C+  C  C-  D+  D  D-  F</td>
</tr>
<tr>
<td>2. Which test format do you believe allowed you to best show what you know on this test? (Check ONLY ONE.)</td>
</tr>
<tr>
<td>_____ Write &amp; Draw  _____ Multiple choice  _____ Pattern Model</td>
</tr>
<tr>
<td>3. How effective do you think were in preparing for this test? (Put an X on the scale below.)</td>
</tr>
<tr>
<td>Not effective preparation 1---2---3---4---5---6---7 effective preparation</td>
</tr>
</tbody>
</table>

163
Test Three Evaluation

NAME: _________________________ Section Number: ____________

Although I need your name for data analysis purposes, YOUR ANSWERS WILL BE KEPT CONFIDENTIAL!

Instructions:
Please tell me how you feel about each part of the exam you just completed. Put an X in one of the 5 spaces to show your agreement or disagreement with each statement.

Part I: Write and Draw

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. This part was difficult.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Part II: Multiple-Choice

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. This part was difficult.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Part III: Pattern Model

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. This part was difficult.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Part IV: Rate your preparation and performance

1. Circle the grade you believe you will receive ON THIS EXAM: (your best guess)
   A A- B+ B B- C+ C C- D+ D D- F

2. Which test format do you believe allowed you to best show what you know on this test? (Check ONLY ONE.)
   _____ Write & Draw   _____ Multiple choice   _____ Pattern Model

3. How effective do you think were in preparing for this test? (Put an X on the scale below.)
   Not effective preparation 1------2------3------4------5------6------7 preparation

164
Final Exam Evaluation

NAME: ________________________  Section Number:________

Although I need your name for data analysis purposes, YOUR ANSWERS WILL BE KEPT CONFIDENTIAL!

Instructions:
Please tell me how you feel about each part of the exam you just completed. Put an X in one of the 5 spaces to show your agreement or disagreement with each statement.

Part I: Write and Draw

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. This part was difficult.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Part II: Multiple-Choice

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. This part was difficult.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Part III: Pattern Model

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Undecided</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I liked this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. This part was difficult.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I felt pleasant (happy, calm) during this part.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Part IV: Rate your preparation and performance

1. Circle the grade you believe you will receive ON THIS EXAM: (your best guess)
   A  A-  B+  B  B-  C+  C  C-  D+  D  D-  F

2. Which test format do you believe allowed you to best show what you know on this test?
   (Check ONLY ONE.)
   _____Write & Draw   _____ Multiple choice   _____Pattern Model

165
3. How effective do you think were in preparing for this test? (Put an X on the scale below.)

Not effective preparation 1---------2---------3---------4---------5---------6---------7 preparation

Very effective

4. How knowledgeable do you consider yourself to be about taking tests at the university level in the U.S. system? (Put an X on the scale below.)

Not knowledgeable 1---------2---------3---------4---------5---------6---------7 knowledgeable

Very knowledgeable

THANK YOU FOR YOUR COOPERATION THIS SEMESTER!
APPENDIX F

TESTING PREFERENCE QUESTIONNAIRE
TESTING PREFERENCE QUESTIONNAIRE

The purpose of this questionnaire is to help English language instructors understand the testing preferences of English as a second language (ESL) students in university courses. This information will help English language instructors prepare ESL students to more effectively and appropriately demonstrate their knowledge on tests. While your help is needed, it is also important to point out that completing this questionnaire is not mandatory. Your answers will be kept confidential. Completing this questionnaire will take approximately 15 minutes.

A. Background Information
Instructions: Please answer the following questions as accurately as possible. Some questions require you to write an answer. Other questions require your putting an X on a scale or in the appropriate blank.

1. Name: ___________________________ 2. Sex: _____Female _____Male
5. Are you a permanent U.S. resident? _____Yes _____No
6. Are you a visiting international student? _____Yes _____No
7. Approximately, how long have you lived in the U.S.?
   _______Years ________Months _______Days
8. Approximately, how long have you attended American universities?
   _______Years ________Months _______Days
9. How knowledgeable do you consider yourself to be about taking tests at the university level in the U.S. system?
   Very knowledgeable _______________________________ Not knowledgeable
   1 2 3 4 5 6 7
10. Which language skill is your STRONGEST/BEST? (Check ONLY ONE.)
    ____Listening ____Speaking ____Reading
    ____Writing ____Vocabulary ____Grammar
11. Which language skill do you believe would help you the most with taking tests in university courses? (Check ONLY ONE.)
    ____Listening ____Speaking ____Reading
    ____Writing ____Vocabulary ____Grammar
12. Have you ever taken an American History class before? _____Yes _____No
    IF YES, HOW MANY CLASSES? ____
**B. Test Preferences**

1. Please rate each of the following test formats based on 10 adjective opposites. Put an X on the number on the seven-point scale for each adjective pair that best represents your attitude toward each test format when taking tests in university courses in English, your second language.

One or more examples of each test format are provided on a overhead to clearly define each type of test format intended. IF YOU HAVE ANY QUESTIONS, please ask Ms. Teemant.

a. True/False Questions:
   only two answers are offered for you to choose among (true/false, yes/no).

<table>
<thead>
<tr>
<th>Adjective</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. difficult</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>b. complicated</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>c. unclear</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>d. boring</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>e. tricky</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>f. unfair</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>g. worthless</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>h. low expectancy</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>i. high anxiety</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>j. feeling</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>k. feeling</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>l. with exam</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Are your attitudes toward true/false questions the same in English as they would be for testing in your native language?

_____ Yes  _____ No  If NO, briefly explain why:

----------------------------------------------------------------------------------

----------------------------------------------------------------------------------

169
b. Multiple-Choice Questions:
you are asked to select among 3 or more possible answers.

<table>
<thead>
<tr>
<th>Option</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. difficult</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>b. complicated</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>c. unclear</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>d. boring</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>e. tricky</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>f. unfair</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>g. worthless</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>h. low expectancy</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>i. high anxiety</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>j. feeling</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>with exam</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
</tbody>
</table>

Are your attitudes toward multiple-choice questions the same in English as they would be for testing in your native language?

_____ Yes  _____ No  If NO, briefly explain why:


c. Short-Answer/Completion Questions:
you are asked to supply a one-word or short-phrase answer.

<table>
<thead>
<tr>
<th>Option</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. difficult</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>b. complicated</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>c. unclear</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>d. boring</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>e. tricky</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>f. unfair</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>g. worthless</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>h. low expectancy</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>i. high anxiety</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>j. feeling</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
<tr>
<td>with exam</td>
<td>1-----2-----3-----4-----5-----6-----7</td>
</tr>
</tbody>
</table>
Are your attitudes toward short-answer/completion questions the same in English as they would be for testing in your native language?

_____ Yes  _____ No  If NO, briefly explain why:

________________________________________________________

________________________________________________________

d. **Restricted-Response Questions:**
give you strict limits to the content and form of your answer. For example, you are asked to list, define, give reasons, etc.

a. difficult 1-------2-------3-------4-------5-------6-------7  easy
b. complicated 1-------2-------3-------4-------5-------6-------7  simple
c. unclear 1-------2-------3-------4-------5-------6-------7  clear
d. boring 1-------2-------3-------4-------5-------6-------7  interesting
e. tricky 1-------2-------3-------4-------5-------6-------7  straightforward
f. unfair 1-------2-------3-------4-------5-------6-------7  fair
g. worthless 1-------2-------3-------4-------5-------6-------7  valuable
h. low expectancy of success 1-------2-------3-------4-------5-------6-------7  high expectancy
i. high anxiety producing 1-------2-------3-------4-------5-------6-------7  low anxiety producing
j. feeling uncomfortable with exam 1-------2-------3-------4-------5-------6-------7  comfortable with exam

Are your attitudes toward restricted-response questions the same in English as they would be for testing in your native language?

_____ Yes  _____ No  If NO, briefly explain why:

________________________________________________________

________________________________________________________

171
e. Extended-Essay Questions:
you are asked to explain, compare, or discuss some issue with almost unlimited freedom in developing the form and scope of your response.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. difficult</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>b. complicated</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>c. unclear</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>d. boring</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>e. tricky</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>f. unfair</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>g. worthless</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>h. low expectancy of success</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>i. high anxiety producing feeling uncomfortable with exam</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

Are your attitudes toward extended-essay questions the same in English as they would be for testing in your native language?

_____ Yes  _____ No  If NO, briefly explain why:

____________________________________________________________________
____________________________________________________________________

2. Which one of the five test formats listed below do you believe allows you to best show what you know on tests?

(CHECK ONLY ONE.)

_____ True/False  _____ Multiple-choice  _____ Short-answer/completion

_____ Restricted-response  _____ Extended-essay
### C. General Test Anxiety

**Instructions:** This scale deals with your feelings about a variety of scholastic situations. Read each statement and decide to what extent it applies to you. Circle the number in the appropriate column under the heading which best describes the **frequency** you experience the feelings described by the statements. Do not ponder the questions, work as rapidly as possible since your first impression is usually accurate. Answer every item.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th>RARELY</th>
<th>ALMOST</th>
<th>ALWAYS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Nervousness while taking an exam or test hinders me from doing well.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. I work most effectively under pressure, as when the task is very important.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. In a course where I have been doing poorly my fear of a bad grade cuts down on my efficiency.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. When I am poorly prepared for an exam or test, I get upset, and do less well than even my restricted knowledge should allow.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. The more important the examination, the less well I seem to do.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. While I may (or may not) be nervous before taking an exam, once I start, I seem to forget to be nervous.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. During exams or tests, I block on questions to which I know the answers, even though I might remember them as soon as the exam is over.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Nervousness while taking a test helps me do better.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. When I start a test, nothing is able to distract me.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. In courses in which the total grade is based on one exam, I seem to do better than other people.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I find that my mind goes blank at the beginning of an exam, and it takes me a few minutes before I can function.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. I look forward to exams.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. I am so tired from worrying about an exam, that I find I almost don't care how well I do by the time I start the test.</td>
<td>1 2 3 4 5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
14. Time pressure on an exam causes me to do worse than the rest of the group under similar conditions.

15. Although "cramming" under pre-examination tension is not effective for most people, I find that if the need arises, I can learn material immediately before an exam, even under considerable pressure, and successfully retain it to use on the exam.

16. I enjoy taking a difficult exam more than an easy one.

17. I find myself reading exam questions without understanding them, and I must go back over them so that they will make sense.

18. The more important the exam or test, the better I seem to do.

19. When I don't do well on a difficult item at the beginning of an exam, it tends to upset me so that I block on even easy questions later on.

THANK YOU FOR YOUR COOPERATION!
APPENDIX G
SAMPLE TEST QUESTIONS FROM AMERICAN HERITAGE 100
<table>
<thead>
<tr>
<th>Instructions: Please follow the instructions in each box.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Draw and explain a true example that, “power struggles may result in cooperation” in Business. (4 points)</strong></td>
</tr>
<tr>
<td><strong>Draw and explain a true example that, “structure may destroy counterpoise” in Education. (4 points)</strong></td>
</tr>
<tr>
<td><strong>Draw and explain a true example that, “power struggles may result in cooperation” in Nature, excluding humans. (4 points)</strong></td>
</tr>
<tr>
<td><strong>Draw and explain a true example that, “structure may maintain counterpoise” in United States History. (4 points)</strong></td>
</tr>
<tr>
<td><strong>Draw and explain a true example that, “power struggles may result in conflict” in Your Own Life. (4 points)</strong></td>
</tr>
<tr>
<td><strong>Draw and explain a true example that, “juxtaposition may result in cooperation” in Your Country. (name your country) (4 points)</strong></td>
</tr>
</tbody>
</table>

What part of the United States political system is more critical to its function than any other part? Explain why. (3 points)
Explain what is meant by "Greek Virtue" (2 points)

Explain what is meant by "Christian Virtue" (2 points)

Draw and explain a true example of "behavior may be influenced by relativity," in Sports. (4 points)

Draw and explain the differences between the American Confederation and the American Federation. (6 points)

American Confederation

American Federation

Do the majority of the people on the earth today behave mainly according to the ideas of "Universalism" or "Relativism?"

Please explain your answer. (4 points)

Define enumerated powers and reserved powers in terms of the United States Constitution. (4 points)

Draw and explain how the present union of the United States can be legally abolished. (4 points)

Draw and explain a true example of "consolidation of power may have positive results" in Your Own Life. (4 points)

The Enlightenment philosophers suggested action out of self-interest rather than virtue was not evil. However, to avoid the dangers of pure self-interest what did they recommend to guide self-interests? (2 points)
Draw and explain a true example of "growth may result in entropy" in United States History. (4 points)

Draw and explain a true example of how the United States government has been beneficial to its citizens. (4 points)

Draw and explain a true example in Your Own Life of how you experienced entropy and the intervention you used to renew growth. (5 points)

Draw and explain a true example of how the United States government has been harmful to its citizens. (4 points)

Write 4 of the major beliefs on the Belief Window of the authors of the United States Constitution. (4 points)

1. 

2. 

3. 

4. 

Draw and explain a true example of an organization of Your Choice that has been, or is, both beneficial and harmful to the people in it, excluding the government of the United States. (6 points)

Note: All the exams will be longer than this one. This practice exam is to help you see the format of the exam. It will also help you on the first exam. No other practice exam will be given. Some exams will include an article to read and a subject pattern model to complete. All exams will have multiple choice questions. You may want to give yourself about 2 to 3 hours to take the first exam to be sure you have enough time to complete it. Any question left blank on exams is marked zero. Your professor will keep completed exams in his office for you to look at until the next exam is given. At that time all exams will be destroyed.
# SUBJECT PATTERN MODEL

(38 points total)

<table>
<thead>
<tr>
<th>Name</th>
<th>Section</th>
<th>Subject pattern</th>
<th>Write the General Principle of your model: (4 points)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(Please print your full name)</td>
<td></td>
</tr>
</tbody>
</table>

Name of your model: ____________________________ (2 points)

On the following lines, briefly describe the pattern of the specific context from which you developed your model: (6 points)

__________________________________________________________________________________________
__________________________________________________________________________________________
__________________________________________________________________________________________

Make a drawing that represents your General Principle. A symbolic drawing is usually better than a literal drawing. (10 points)

__________________________________________________________________________________________

<table>
<thead>
<tr>
<th>Write and Draw a true, specific example of your General Principle in Nature (plants/animals) excluding humans: (4 points)</th>
<th>Write and Draw a true, specific example of your General Principle in United States History: (4 points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Write and Draw a true, specific example of your General Principle in Your Personal Life, not your mother, uncle, sister, etc.: (4 points)</td>
<td>Write and Draw a true, specific example of your General Principle in a context of Your Choice, excluding Nature, US History, and Your Personal Life: (4 points)</td>
</tr>
</tbody>
</table>

179
APPENDIX H

EXAMPLE TRANSCRIPT: SUBJECT 08
NOTE: All interviews were transcribed as they were actually spoken. Excerpts from any
of the transcribed interviews presented in Chapters 4 or 5 have been edited.

FACE CARD ID: 43582 Interview: 08 Interviewer: Annela Teemant
Date: Nov. 11, 1996
08 OSU Architecture Undergraduate Male Spanish Puerto Rico
INTERVIEW NO.: 08 Question No.: 1 Card ID: 43981

Data:
T: I already know your name is _________. Is that how you say it?
S: Yeah.
T: What country are you from?
S: I'm from Puerto Rico.

INTERVIEW NO.: 08 Question No.: 2 Card ID: 44172

Data:
T: Okay. how long have you been in the United States?
S: This is my second year.
T: Only second?
S: Yeah.
T: Okay. Good.
T: How long have you been at OSU?
S: Two years.
T: Two years?
S: Two years.
T: Okay. Did you go to a university in Puerto Rico, did you study there?
S: Yeah, I studied in Puerto Rico like just for one year, but I didn't like it. It's not like a challenge for me. You know, it's like everything is Spanish, and I understand everything. I think here is more like a challenge for me. And I like it. I like to learn new stuff, to meet a lot of people. Because in Puerto Rico you are only going to see Puerto Rican people. You never see like Indians, Chinese people, or anything like that. So-

T: Okay. What are you planning on majoring in?
S: Architecture.
T: Okay. Are you already starting those architecture courses yet?
S: Yeah, I'm already taking some history of architecture, but I need to apply to the school for next year.
T: To get in.
S: I heard too like choose xxx classes so-
T: Okay. Good. Well, like I told you on the phone, and when I came to the class, I'm interested in your experiences, and the concerns or worries you've had in testing situations since being here in the U.S. Trying to understand better what English as a second language students worry about or work with. So I was wondering if you remember the first test you took at the university, and if you could describe that experience.

S: Yeah, it was like math exam. And the problem that I had was when I took those like verbal problems. Then, for example, like sometimes you don't understand what they mean like if like some direction that they give you, and it your first exam, and you know you just block. And you don't know. That was my first experience, and most of the test was fill blanks, and match some answers, but the real problem was the verbal. And the other problem is when you have to make an essay. For, for example, my architecture classes, and you have to put some of your point of view or what you think, you know like at the first you try to memorize everything from the book, you know. Like okay if they ask this I can write the saying of the book, but when they ask you like "what you think about that" you know, it's kind of hard.
T: Okay. Well, those are good, you're giving me good information already. Ahoy, if you were to say what the major differences are between the testing system in the United States versus the testing system you're used to in Puerto Rico, what would you say the differences are?

S: Actually, it's pretty much the same.

T: Yeah, because Puerto Rico is attached.

S: Yeah, because Puerto Rico is attached. But it's more like, always is based on the books, it's weird if they ask you like what you think. It's more like you have to answer this because you read that. You know, it's like more memorizing. It's not like a lot of thinking, you know like for example in the English that I'm taking right now, we we read this essay, and you have here to be yourself, and you have to do what you think, or and there is you do what the instructor say, not what you think.

T: Okay. Can you think of any other main differences between the system.

S: Main differences. The math is the same. And I never took like any architecture classes, so-

T: So you wouldn't know.

S: No.
T: Okay. When you are taking a test in English here, what are the main problems that you encounter when taking a test in math, or architecture because of language?
S: Sometimes they use some words that I never hear about it, some difficult word for me to understand, and I have to read a lot of time to understand what they want to say. You know. Like try to find the meaning about the context of the sentence. And what other problem?

T: What have you learned about taking tests here in the United States? Maybe that's another way to think about it.
S: Ah, to take. First, like you have a little time to do an exam, and you have to be prepared completed to answer the question for example in the in the exam I took for English. They give like one hour and forty minutes, but you have to think, to organize, and to check the grammar. You know, and sometime for example Puerto Rico the you say first the adverb and then the verb. You know it's kind of different.
T: Right.
S: And then you have to rephrase those because you first you thinking Spanish, and then you want to rephrase that in English. And it is kind of hard because I don't have enough time to do that. And another thing will maybe be like
T: You said that you really have to know everything.
S: Yeah.
T: What do you mean by that?
S: Like I have to understand each word because maybe I think maybe this an important word. If I don't know, maybe I'm going to answer something wrong, and they don't ask about that. You know, something I I have to, maybe it's me but

T: No, yeah, no.

S: I have to.

T: So when you say you have to know everything, you're talking about the language?

S: Yeah.

T: You're not talking necessarily about what the test question's about?

S: Yeah, both.

T: Both.

S: For example, in architecture, you have to memorize a lot of periods, and a lot of stuff.

For example, compare he want to comparison from one period to other period. And, you have to know everything, you know, like. It's- I don't know. It's kind of hard because he put a lot of stuff that you don't even think that was important, you know.

T: And now it is?

S: Yeah. I know for the next.

T: And you have to memorize.

S: Um hum. You have to memorize a lot.
T: Do you have any strategies that you use when studying or taking tests?
S: I try to read everything, like before class to know what they are talking about. Because sometime if I'm going to classes I don't read before, I don't know where the hell are they talking about. So, I try to read before class, and do like summaries, or the most important, and take a lot of notes.
T: So you write notes out?
S: Yeah.
T: In class you mean, or from the readings?
S: In class, and then of the reading too.
T: You do both.
S: Yeah.
T: Okay. When you get the test, let's say it's an architecture test, is there anything you-strategy you use while taking the test to help yourself take the test?
S: I first read everything like all the test.
T: Okay.
S: And then I try to find in if they have for example if he make some question in the next page he gives the answer, you know he's talking about the same. You know, to have some idea if I've doing right. Um. At- but- at the first I did what I'm sure that I know quickly, and then I try to double check, and if I don't know then I try to guessed.
T: Well, the best you can.
S: Yeah.
T: Any strategies you're using now that you are using now after two years that you didn't know about at first, when you first got here?
S: Yeah, actually, reading a lot. Reading a lot because when you read you know the order of the sentence. You learn a lot about grammar, and your vocabulary. And I read all that they say because it's almost what are you going to see in the exam, so that's pretty much what I do.
T: So you read more now than you used to?
S: Yeah, yeah.
T: At first, you figured out reading was important.
S: Yeah. Because if I read in my home, if I don't know some word, I can check in the dictionary. Because I'm not carrying my dictionary everywhere, so in my home I can check a lot of word I don't know.
T: Do you use your dictionary in test situations?
S: I can't.
T: They don't let you.
S: No, they don't let me.
T: Okay.
S: So.
INTERVIEW NO.: 08    Question No.: 8    Card ID: 46482

Data:

T: Ah, yeah. Of the grades you've gotten on tests since you've been here, do you think those grades reflect what you really know? Are they accurate?
S: No. Sometime they don't reflect what you know. Because maybe at the first I remember my first exam. I wa, I knew everything because I studied pretty hard. You know, first exam you want to do okay.
T: Yes.
S: But when I went to the the exam, and I read those verbal problem that I didn't understand, you know, I couldn't do that. Because I don't know what they were asking to do.
T: Right.
S: So, I don't think that was like wasn't good, because I didn't prove what I know, you know.
T: In math, especially with the word problems.
S: Yeah, math is more verbal problem.
T: Any other class that you have felt that way where your grade didn't really show what you knew?
S: In my first architecture class, when I had to write like two essays, and they give you like fifty minutes, and you have to write two essays, and you have to matching, and vocabulary and a lot of stuff. You know. And at the first, I tried to do an essay, but it was like really really poor, you know. Like my vocabulary. My vocabulary was not like the other people. They know a lot of word, and it was really really poor. So, but, I knew it, the the what they were asking.
T: Oh, you knew what they were asking, but you couldn't show it it?
S: Show it. Yeah.

T: And you say it's because of the vocabulary?

S: Yeah, sometimes vocabulary, and other time is because I didn't ask what they were asking.

INTERVIEW NO.: 08  Question No.: 9  Card ID: 46718

Data:

T: Okay. Um. When, When you take tests, what kinds of tests would you prefer to take?

S: Ah verbal, like some American talking to me.

T: Talking?

S: Because if he don't understand what I'm saying, I can try to explain to him. You know, because if you have to write, they are going to grade you about what you write, not what you really think. Because we can debate, what you think, or what I think. Why it is right or why it is wrong.

T: Have you ever taken any verbal tests?

S: No.

T: Did you take some in Puerto Rico?

S: Yeah. For one psychology class that I took.

T: You got to take an oral test?

S: Yeah.

T: So you think you would do better if you could talk your answers through rather than write?

S: Yeah.
T: Okay. Um. I have a list of types of tests. For example, I'll let you hold that for me. Um, usually we talk about written tests, not an oral test isn't here, but we have you know true/false questions, multiple choice questions, short answer--you know where you put in the missing word or the answer is President Clinton or something--and then we describe this as restricted response, but those are like short essays, where you list or define, or give reasons--and then, of course, the extended essay. Where like you said, fifty minutes, two essays. When you're taking a test, which one would you prefer?

S: Multiple choice.

T: Really?

S: Yeah.

T: Tell me why.

S: Why? Because you have the answer in the paper. It's no way that you have to write or think, I have to do that it just it's right here all that you have to know is just remember the answer. You know. It's just like more memorize, like, it's pretty easier.

T: Okay. Would you say that's the easiest kind of test to take in English for you?

S: This and true or false because it's the same, it's just true or false. You know, if you can understand what they are saying you can know if that true or false. You know. This is pretty much the same. It's just-

T: Short answer.

S: Yeah. It's must memorize.
T: Okay. Which is the most difficult in English when you are taking a test in English, which is the hardest for you?
S: Essay.
T: Essay? Tell me why.
S: Because you have to organize and develop a topic. And you have to have an order, and to put your ideas in order. And also, you have to be as specific about what are they asking because you don't have a lot of time. And if you don't know, for example, some word, you have to change the whole sentence because it doesn't make sense. And you lose a lot of time.
T: Time.
S: You know, you have to use words that you know, and sometime it's not like the best word you know. It's- if you are doing an essay for an exam, you have to have a good vocabulary you know to impress your teacher because you are not going to use like pretty, or beautiful, you know, those are some word that everybody use. It's not like some college student should be use. You know what I'm saying?

T: Yes. Um huh. Okay. If you, ah were taking a test in Spanish, which would be the easiest for ya?
S: Um, I think any-
T: Any of them, no problem?
S: Yeah. No problem.
T: Okay. Because it's your native language?
S: Yeah, it's my native language.
T: Which one's the hardest for you in Spanish? When you take tests at home.
S: Sometime true or false, because they use some kind of word that is kind tricky. That you don't know if they are true or false. For example, the true or false that I took, you have to explain why it's false. So-
T: That makes it harder?
S: Yeah. But I think in the Spanish it is pretty nice if I can do an essay because I can put everything that I know, and maybe I can add something that I think, you know two extra points. Or if I don't know the answer, I can write something pretty much like what what are they asking. But in English, I can guess, but it's not the same.
T: It's harder for you.
S: Yeah, it's harder for you.

INTERVIEW NO.: 08 Question No.: 10 Card ID: 47621

Data:
T: Okay. That's interesting. Good. Okay. Um, I was gonna ask if which language skill do you think helps you the most when you are taking a test, or if you could improve any language skill--reading, writing, listening, speaking, grammar, vocabulary, pronunciation--which one do you think would help you the most in a testing situation?
S: I think grammar and vocabulary. I think those are so important because the instructor, she gave us the exam today back-
T: Yes.
S: and I had like C, and she told me that I had good ideas, and well organized, but my grammar was poor. And sometime I put the verb before the noun or after the noun. You know, some verb agreement and statement agreement. I think that that kind of help me a lot.
T: If you're taking a test in architecture or another subject, would it be the same?
S: In architecture is like actually pronunciation. Because you have to do a lot of presentation in architecture. If you do a model, then you have to present, and explain why you did that, and the same the portfolio, and you have to explain everything. And sometime the pronunciation make make it difficult, because they say "What? Why?" And you know, the the instructor here are from everywhere, you know. And everybody have their own accent. And it's kind of difficult to you know, for example the word what you pronounce completely different. For example, Puerto Rico you know there are xxx is the same word, but the pronunciation is completely different.

T: That's right.

S: You know, and I think that can help a lot too.

INTERVIEW NO.: 08 Question No.: 11 Card ID: 47909

Data:

T: Can help. Do you think there is anything that professors can do to help you as a second language student. I mean English is your second language, is there anything professors can do in your classes to make testing better for you?

S: Um, I think that I should be take the same exam that everybody took, but I think that if I'm doing like a pretty good job in in the class, and for example I fail the exam, they should be call me when, well actually I should be go to them, and explain to them, and just have a talk, and maybe he can ask me about the question that I have wrong. And maybe he can see that I know the the material. It's not like I don't I don't under- I didn't study or something like that.

T: Right. Anything else you think that they could do to make help you do better on tests in their class?
S: Um. It's all depend of the instructor. Because some instructors speak so fast, and you can't hear him, and what are they saying. And other like prefer more from the books because you are going for example to some class and they talk and talk, but they never say anything about the book. And when are you going to exam, you are not going, you don't know if you are going to say some stuff from the book, or whatever he said, you know.

To be a more specific on-

T: On, on wh-, where what material is covered in the test?

S: Yeah. Yeah.

T: Okay. Anything else?

S: No, I think.

INTERVIEW NO.: 08  Question No.: 12  Card ID: 48204

Data:

T: That's about it. Did you study in an intensive English program before coming to the United States?

S: No.

T: And when you came here, you came directly to the university?

S: Yeah, directly to university. I never stay here to learn English or anything like that.

T: Okay. Okay.

S: I took English class on high school, but the teacher was like she was not the best. So, um, when you take classes on high school you never think you are going to study on United States, so you never pay attention what they are saying. It's more like just memorize some verbs, and some xxx. For example, I think here is like elementary school, some textbook like Mom, Dad, you know. It's not like, yeah it's pretty easy.
T: Is there anything that you've wanted to tell me about testing that I haven't asked you yet? Anything you can think of?

S: Ah, for example, in the exam we took, I think it not fair like the way that they grade the exam. Because I don't think they should be expect that you do like an essay like native American people. You know, they can't expect like you have perfect grammar, and I think that I have to have the test xxx and all that stuff.

T: Say that again, you have to have the-

S: The thesis statement and the introduction.

T: Oh, thesis statement. Oh. sorry.

S: Thesis statement. Introductory paragraph and all of that, but I I have all of that but they told me my problem was the grammar. What I I'm not expecting that they give me A, but I think that they can't expect that you have like a perfect grammar sentence you know like verb agreement, everything like that. Because sometime you can't memorize some example, or you can read this is the the rules to do that but not every time you can follow the rules. You know. They have a lot of exceptions that I don't know.

T: Right.

S: But, I think

T: Do you have that same problem in your architecture class, with the essay exam?

S: Yeah.

T: Did they grade the grammar?

S: No, in architecture no. But I don't have the they make some marks what you want to say about that, and I have to go and say but it's pretty much from English class. Like they xxx.
T: They focus on that a lot.

S: I think it's important

T: Sure.

S: Because if I'm going to stay here, if I'm going to send a letter to somebody, I-

T: You want it to be right.

S: Yeah.

T: Okay, good. Anything else? No? Well you passed my test.
APPENDIX I

TEST FORMAT EXAMPLES
A. TRUE/FALSE QUESTIONS

FOR EXAMPLE IN A COMPUTER CLASS:
T   F   The spreadsheet concept comes from Michigan University.

B. MULTIPLE-CHOICE QUESTIONS

FOR EXAMPLE IN A CHEMISTRY CLASS:
Gases show significant deviation from the gas laws at
a) low pressure and high temperature
b) high pressure and low temperature
c) greater average kinetic energy
d) low volume and high temperature

C. SHORT-ANSWER/COMPLETION

FOR EXAMPLE IN A HISTORY CLASS:
Which president of the United States signed the Panama Canal Treaty?

FOR EXAMPLE IN A STATISTICS CLASS:
The __________ is the best measure of central tendency when you are
interested in peak of the distribution.

D. RESTRICTED RESPONSE (list, define, give reasons, etc.)

FOR EXAMPLE IN A SCIENCE CLASS:
List the environmental conditions that affect the color of leaves in the autumn.

What is the relationship between photosynthesis and the appearance of red leaves?

FOR EXAMPLE IN A SPEECH CLASS:
Define and give an example of personification.

FOR EXAMPLE IN A LANDSCAPING CLASS:
Give two causes for the revolution in landscape gardening in the eighteenth
century.

E. EXTENDED ESSAY

FOR EXAMPLE IN A HISTORY CLASS:
Write an essay comparing and contrasting the China's concept of government -- its
purposes, its functions, the role it plays in citizens' lives -- with the American
concept of government.

FOR EXAMPLE IN A SHORT STORIES CLASS:
When Isaac Singer was given the Nobel Prize for Literature in 1978, he was
praised for 'bringing universal conditions to life.' What "universal conditions"
does he bring to life in "Naftali the Storyteller and His Horse, Sus"?
APPENDIX J

REASONS FOR ATTITUDES TOWARD TEST FORMATS

A. True/False

01 I would be able to understand the teacher's way of thinking. Sometimes those questions get really tricky if I do not see things the way the teacher does.

03 I've noticed that sometimes true/false questions were based on little details, and it's hard to get little details when we're taking notes for example.

06 I would feel a little boring.

07 Because I can get what the sentence says better in Arabic.

08 It's more simple in French, but it's more challenging in English.

09 The wording in my language would be easier to understand, so I could know exactly what is wanted.

10 The words sometimes are tricky, so I understand something different.

16 Because in my own language vocabulary is easier.

20 Because when American teachers try to be tricky with words, I get confused. Because no matter how good I am in English, I won't be as good as native speakers.

21 We have not very many tricky or unclear questions in my native language exams.

23 A little more comfortable in own language.

26 Because sometimes the wording can be unclear double meaning -- you have to try to understand what the professor has in mind.

27 The wording of the true/false questions may be deceiving.

29 In my country I can understand fully the meaning of the words in the often tricky questions.

33 Because are simple word in English may make very different the meaning of a sentence and because I don't understand this language very well.
B. Multiple-Choice

03 Same thing as true/false questions [i.e. I've noticed that sometimes true/false questions were based on little details, and its hard to get little details when we're taking notes for example.] And also they like to make it tricky and play a little bit with the language which makes it harder.

04 In English they could be trickier because of vocabulary you're not familiar with, or different ways of expression.

05 No because I have a better knowledge of the vocabulary and words that can make a difference in the meaning of a question.

06 I could be able to understand the wording in Japanese better.

10 Sometimes the answer that we have to select are very similar to the things we study in class or that we read in the book, so if I understood the main idea of the subject, I won't get the answer right.

20 When the teachers try to be tricky with words, I get confused.

21 Sometimes I find vocab I don't know in answers in English exams, and it makes difficult to give right answers.

25 Vocabulary = I need to know the meaning.

26 Because I can understand written English.

27 Again, the wording may be deceiving, and the answers may be tricky.

29 Same than for true or false. [In my country I can understand fully the meaning of the words in the often tricky questions.]

30 I would never have to take an exam with options from A through I, being all answers correct and we have to choose the most correct. That's stupid.
C. **Short-Answer/Completion**

02 Because I don't know lots of words that could fit the blank as an English speaker could.

04 Many times you're asked for specific words and it may happen that you can't remember that one word in English, but you do remember it in your native tongue.

10 Sometimes when I write my answers the teacher would have difficult time understanding what I mean on my answer.

11 In the English language, words might be stated in a different, so even though you know the material very well, you might not answer the question correctly.

14 I can express myself better in French.

16 No, because I would do better in my language.

21 Because of my poor language skills, I cannot remember the right word even I understand a questions.

25 Spelling. Have to remember the word in English.

26 The problem is not the language but understanding what answer the professor has in mind.

29 Because is easier for me to learn a whole concept in my language. I understand usually the concepts, but for this kind of question everything is centered in a single word that you can easily forget.

35 Understanding sometimes what the teacher wants exactly.
D. Restricted-Response

01 If I did it in Spanish, I would have all the words to explain what I think.

02 Because I don't know how to write very well in English as I do in Spanish. Well, not yet.

04 Again, vocabulary. Many times you can explain perfectly in your own tongue, but can't put the words together to the exact meanings of what you want to say.

05 No, because in my language sometimes I can memorize and relate better because I have no limits in my expression and I have a better knowledge of the language.

12 Multiple choice is enough to evaluate students.

18 Sometimes if I cannot use my dictionary while taking the exam, I cannot express myself clearly enough as when I would write it in my language. I may know the answer but I just don't know how to say/write that. But I like it.

21 It takes time to answer. I don't like to have time limit for this kind of questions.

23 I haven't had experience here yet.

24 Language limitation.

25 Sometimes I forget the right names.

26 The problem is how well prepared I'm for the exam.

28 Probably no, because it would be more easy for me to explain my ideas in my own language.

32 Vocabulary can sometimes be a problem when something has to be explained short and concise.
E. Extended-Essay

01 My vocabulary is very limited.

02 Because I still have problems with my English writing.

03 Our style is not as developed in English and some grammar structures may be wrong which gives not so good of an aspect to the essay.

04 Same as before. [Again, vocabulary. Many times you can explain perfectly in your own tongue, but can't put the words together to the exact meanings of what you want to say.]

06 I always worry about grammar and vocabulary when I have an essay test. Sometimes I have hard time expressing what I really want to say.

09 Points will be deducted for spelling and grammar errors, so if we don't do these things correctly we get poor grades. Unfair.

11 Again, the material might be known but the question could be stated differently.

18 Please see answer in question D: [Sometimes if I cannot use my dictionary while taking the exam, I cannot express myself clearly enough as when I would write it in my language. I may know the answer, but I just don't know how to say/write that. But I like it.]

21 The same reason as D: [It takes time to answer. I don't like to have time limit for this kind of questions.]

24 Language limitation.

26 I need more time to write in English.

28 Because it easy for me to do an essay in Spanish.

30 Because of the native language.

31 I think in some things I better express myself in Spanish, and others in English.

32 It can sometimes be hard to make yourself clear, but it is anyway the testing form that is most challenging and interesting.

33 Because in my language, I could feel free to express even with more words.

35 Understanding.
REFERENCES


210


211


212

