Portland State University

PDXScholar

Dissertations and Theses

Dissertations and Theses

4-28-1995

An Examination of the English Vocabulary Knowledge of Adult English-for-academic-purposes Students: Correlation with English Second-language Proficiency and the Validity of Yes/No Vocabulary Tests

Robert Scott Fetter
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds

Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Fetter, Robert Scott, "An Examination of the English Vocabulary Knowledge of Adult English-for-academic-purposes Students: Correlation with English Second-language Proficiency and the Validity of Yes/No Vocabulary Tests" (1995). *Dissertations and Theses*. Paper 4903.

<https://doi.org/10.15760/etd.6779>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

THESIS APPROVAL

The abstract and thesis of Robert Scott Fetter for the Master of Arts in Teaching English to Speakers of Other Languages were presented April 28, 1995 and accepted by the thesis committee and the department.

COMMITTEE APPROVALS:

Kimberley Brown, Chair

Jeanette DeCarrico

Thomas Dieterich

Sandra Wilde
Representative of the Office of
Graduate Studies

DEPARTMENT APPROVAL:

Beatrice Oshika, Chair
Department of Applied Linguistics

ACCEPTED FOR PORTLAND STATE UNIVERSITY BY THE LIBRARY

by

on 17 July 1995

ABSTRACT

An abstract of the thesis of Robert Scott Fetter for the Master of Arts in Teaching English to Speakers of Other Languages presented April 28, 1995.

Title: An examination of the English vocabulary knowledge of adult English-for-academic-purposes students: Correlation with English second-language proficiency and the validity of YES/NO vocabulary tests.

The importance given to vocabulary in second language instruction and in theories of second language acquisition has increased greatly in the last fifteen years. It is thus important for second language teachers/researchers to have valid, useful methods of assessing the vocabulary needs and vocabulary knowledge of their students as well as valid and useful methods of assessing the efficacy of various methods and techniques of teaching and learning vocabulary.

This study examines the usefulness and validity of a relatively new type of checklist vocabulary test method known as the 'YES/NO' method. In the YES/NO method, nonsense words are listed together with real test words. A subject's test score is calculated by applying both the percentage of real

words checked and the percentage of nonsense words checked to a mathematical formula.

Sixty-six students enrolled in a college-level English-for-academic-purposes (EAP) program took three vocabulary tests. Correlation was calculated between the participants' scores on the 'Structure' and 'Listening' subsections of the Comprehensive English Language Test (CELT) and the participants' scores on all three vocabulary tests scored both as YES/NO tests and as simple-checklist tests.

The following three findings were noteworthy: (1) correlation between CELT subtest scores and vocabulary test scores was more consistent and stronger when the vocabulary tests were scored as simple-checklist tests compared to when they were scored as YES/NO tests, (2) few students scored above recommended exit-level scores on the CELT subtests but below 5,000 on the vocabulary tests, and (3) a 120-real-word vocabulary test correlated more consistently and strongly with the CELT subtests than either of two 60-real-word vocabulary tests.

Three conclusions were made: (1) adult EAP students preparing to study at English-medium institutions of higher education need knowledge of the 5,000 most-frequent words [lemmas] of English, (2) the YES/NO method of testing the L2 vocabulary knowledge of adult EAP students is not better than the simple-checklist method, and (3) a good direction for work on the improvement of tests intended to measure the L2

vocabulary knowledge of adult EAP students may be to explore how to elicit valid responses on long simple-checklist tests.

AN EXAMINATION OF THE ENGLISH VOCABULARY KNOWLEDGE OF
ADULT ENGLISH-FOR-ACADEMIC-PURPOSES STUDENTS:
CORRELATION WITH ENGLISH SECOND-LANGUAGE PROFICIENCY
AND THE VALIDITY OF YES/NO VOCABULARY TESTS

by

ROBERT SCOTT FETTER

A thesis submitted in partial fulfillment of the
requirements for the degree of

MASTER OF ARTS
in
TEACHING ENGLISH TO SPEAKERS OF OTHER LANGUAGES

Portland State University
1995

ACKNOWLEDGEMENTS

I would like to thank the members of my committee: Kimberley Brown, Jeanette DeCarrico, Thomas Dieterich, and Sandra Wilde. Special thanks are owed to Kimberley Brown for her kind advice, encouragement and timely reading of numerous drafts.

I would also like to acknowledge the invaluable support I received from my parents, family--and most of all--my wife, Mika.

TABLE OF CONTENTS

	PAGE
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	vi
LIST OF FIGURES.	viii
CHAPTER	
I INTRODUCTION	1
Questions and Hypotheses	6
Research Question 1	
Hypothesis 1	
Hypothesis 2	
Research Question 2	
Hypothesis 3	
Research Question 3	
Hypothesis 4	
II REVIEW OF LITERATURE	9
L2 Abilities Needed by EAP Students	9
Defining and Counting Words	10
L2 Vocabulary Knowledge and EAP Ability	12
Reading	
Writing	
Listening	
Vocabulary Tests.	18
YES/NO Vocabulary Tests	23
Overview of the YES/NO Test	
Procedure	
Development, Use, and Validation	
of YES/NO vocabulary tests	

	Rationale for the Present Study	33
III	METHOD	35
	Subjects.	35
	Instruments	36
	Procedures.	38
	Hypotheses and Analyses	40
	Test Scores	
	Hypothesis 1	
	Test of Hypothesis 1	
	Hypothesis 2	
	Test of Hypothesis 2	
	Hypothesis 3	
	Test of Hypothesis 3	
	Hypothesis 4	
	Test of Hypothesis 4	
IV	RESULTS.	44
V	DISCUSSION	56
	Implications and Suggestions for	
	Future Research	63
	Conclusions	72
	REFERENCES	73
	APPENDICES	78
A	TEST A	78
B	TEST B	80
C	RANK OF TEST WORDS	82
D	VOCABULARY TEST INSTRUCTIONS	86
E	CORRELATION MATRIX FOR ENTIRE GROUP.	88
F	DATA FOR ALL SUBJECTS.	90
G	CORRELATION MATRIX FOR EDITED GROUP.	94

H	CORRELATION MATRIX FOR LOW FALSE ALARM RATE GROUP	97
I	CORRELATION MATRIX FOR HIGH FALSE ALARM RATE GROUP	98
J	NEW CHECKLIST TEST	101

LIST OF TABLES

TABLE		PAGE
I	Lexical Coverage Provided by the Approximately 3,000 Most- Frequent Word Families of English.	17
II	95% Confidence Intervals for a 50% Score on Various Size Vocabulary Tests Sampled from a 5,000-Word List	21
III	Correlation Between YES/NO Tests and CELT Subtest Scores from Meara et al. (1994).	32
IV	Correlation Among All Three CELT Subtests and All Three Vocabulary Tests for All Subjects	45
V	Correlation Among CELT Subtests and Vocabulary Tests for All Subjects Except Five Subjects with Very High False Alarm Rates	48
VI	Distribution of Test C False Alarm Rates for All Subjects	50

VII	Correlation Among CELT Subtest Scores and Vocabulary Test Scores for Subjects with Test C False Alarm Rates Less Than or Equal to 10%	.51
VIII	Correlation Among CELT Subtest Scores and Vocabulary Test Scores for Subjects with Test C False Alarm Rates Greater Than 10%	51

LIST OF FIGURES

FIGURE	PAGE
1. Correlation between scores on Test C scored as a simple-checklist test and CELT-L/S scores for all subjects.	54
2. Correlation between CELT Listening subtest scores and scores on Test C scored as a simple checklist test for all subjects.	55

CHAPTER I

INTRODUCTION

This study examines the validity and usefulness of a relatively new type of test designed to measure the vocabulary knowledge of adult English-for-academic-purposes (EAP) students. This relatively new type of test--known as a 'YES/NO' test--is, paradoxically, a variation of one of the oldest types of test procedures used to measure vocabulary knowledge--the checklist procedure (Read, 1988). In the most common type of checklist test, the 'simple-checklist' test, the test words are listed on paper, and subjects mark the words that they know.

Simple-checklist tests have been used widely in studies of first-language (L1) vocabulary size (Goulden et al., 1990). In such studies the number of test words must be very large. This makes the multiple-choice test procedure impractical as it requires too much response time per item. Thus, simple-checklist tests are attractive because they allow a large number of words to be tested in a short period of time. However, when simple-checklist tests are used, there is no way to tell whether or not subjects have misrepresented their knowledge. Further, it has been widely believed that subjects will tend to overestimate their

vocabulary knowledge on simple-checklist tests (Read, 1988). The YES/NO test procedure was created by L1 researchers to account for these problems.

A YES/NO test is a checklist test which lists some 'nonsense' words together with the real test words. Nonsense words are words, created by the test maker, that look like real words but are not. 'Deptle,' 'consile,' and 'intone' are examples of nonsense words. In theory, as the percentage of nonsense words that a subject marks rises so does the number of real words that are marked but not truly known by the subject. The percentage of nonsense words that a subject marks and the percentage of real words that a subject marks are applied to a mathematical formula to calculate a subject's score.

However, both the L1 and second-language (L2) studies that suggest validity for the YES/NO test procedure have only compared YES/NO vocabulary tests with multiple-choice vocabulary tests. The claim that YES/NO vocabulary tests are a more valid method than simple-checklist tests to measure the vocabulary knowledge of adult L2 learners has not, to the best of my knowledge, ever been tested. This thesis aims to conduct such a test.

Vocabulary test procedures are important to study as vocabulary tests can be useful in several ways. First, vocabulary tests can help in setting targets for instruction by helping to describe the vocabulary knowledge of students

who have successfully acquired a certain level and type of L2 ability. Second, vocabulary tests can help determine where specific students' vocabulary knowledge stands in relation to specified target levels. Third, vocabulary tests can help to measure growth in L2 vocabulary knowledge over time under different conditions such as different teaching methods. Thus, YES/NO tests may be useful as vocabulary tests that can help in learning more about how many and which words students need to know, how many and which words specific students do know, and how different students are acquiring their L2 vocabulary knowledge.

Vocabulary knowledge is an important aspect of L2 competence to study. Several researchers have recently claimed, for example, that for the acquisition of the L2 abilities most needed by adult EAP students--reading, listening, and writing (Sloan, 1994)--vocabulary knowledge is the most urgent need (Fox, 1979; Kelly, 1991; Laufer, 1992; Laufer & Sim, 1985; Strother & Ulijn, 1987; Ulijn, 1984). Other researchers have claimed that EAP students need knowledge of the most-frequent 5,000 word families of English in order to be successful readers and listeners in academic contexts (Kyongho & Nation, 1989; Laufer, 1992, Ostyn & Godin, 1985). These claims have important ramifications for EAP instruction. However, much of the evidence for these claims has been indirect; there is actually little direct evidence of how many and which words different groups of EAP

students know at various levels of proficiency. For example, the claim that EAP readers need knowledge of the most-frequent 5,000 word families of English is based on the combination of evidence that knowledge of at least 95% of the word families in a given text is required to read that text with confident comprehension and without frustration and evidence that it takes approximately the most-frequent 5,000 word families of English to provide greater than 95% lexical knowledge of most written English texts. While these two bodies of research combine to suggest that EAP students need knowledge of the most-frequent 5,000 word families of English, there is little direct evidence of the vocabulary knowledge of EAP students in terms of number of word families known. More direct evidence of the vocabulary knowledge of specific groups of EAP students might be useful.

In order to conduct this type of research, new types of vocabulary tests are needed. Not only are new procedures for testing word knowledge required, but new methods for choosing test words also need to be employed. Most previous research on vocabulary knowledge has been based on the concept of 'graphic words'. Graphic words are unique strings of letters surrounded by space. It has recently been suggested, however, that the construct 'word family' may better reflect the organization of the mental lexicon and be a useful construct in vocabulary research (Carver, 1994; Read, 1988). A 'word family' is a group of related graphic words such that

knowledge of any member of the family helps in acquiring knowledge of any other member.

In order to test EAP students' knowledge of the most-frequent 5,000 word families of English, a test would need to be created by sampling words from a list of the most-frequent 5,000 word families of English. No tests of this type exist. Moreover, word-family frequency lists are not common. To the best of my knowledge, there is only one word-family frequency list of American English that includes the most-frequent 5,000 word families of English--Francis and Kucera's (1982) 'lemma-frequency list'. A lemma consists of a base word and its inflectional variants that belong to the same part of speech. The lemma is a somewhat conservative operational definition of 'word family,' but it is a good practical choice since word-frequency lists based on other operational definitions do not exist. Francis and Kucera's list is a product of work with the 'Brown Corpus' which began in 1961. It is a unique list that may be useful in testing the vocabulary knowledge of EAP students.

In this study three vocabulary tests will be created by sampling words from Francis and Kucera's (1982) lemma-frequency list. Then, EAP students' knowledge of those words will be tested using the YES/NO procedure and the simple-checklist procedure. Correlation among scores on the three vocabulary tests and the 'Structure' and 'Listening' subtests of the Comprehensive English Language Test (CELT) will be

examined. By using this method, it is hoped that some knowledge can be gained concerning the validity and usefulness of the YES/NO test procedure. It is also hoped that some knowledge can be gained concerning the L2 vocabulary knowledge of adult EAP students and the relationship between this knowledge and the level and type of L2 proficiency needed by adult EAP students.

QUESTIONS AND HYPOTHESES

Research Question 1

Is the YES/NO test procedure a more valid measure of the L2 vocabulary knowledge of adults than the simple-checklist procedure?

Hypothesis 1

Scores on Test A, Test B, and Test C will correlate more strongly and more consistently with CELT Structure subtest scores and CELT Listening subtest scores when the vocabulary tests are scored using the YES/NO procedure than when scored as simple-checklist tests.

This hypothesis is based on the assumption that there is a strong correlation between vocabulary knowledge and L2 proficiency (Gass & Selinker, 1994). It is assumed that scores on the more valid test of vocabulary knowledge will show higher correlation with scores on listening and grammar tests.

The terms 'more strongly' and 'more consistently' are used here as the interpretation of the differences among correlation coefficients in a correlation matrix is entirely subjective.

Hypothesis 2

The correlation between checklist vocabulary test scores and CELT subtest scores for subjects with 'high' False Alarm Rates (greater than 10%) versus subjects with low False Alarm Rates (less than or equal to 10%) will be meaningfully different for simple-checklist scores but not for YES/NO scores.

According to YES/NO-test theory, as subjects mark more nonsense words they will also mark more real words that they do not really know. The theory claims that the YES/NO procedure properly accounts for this but that the simple-checklist procedure does not.

The term 'meaningfully different' is used here as the interpretation of the differences among correlation coefficients in a correlation matrix is entirely subjective.

Research Question 2

What is the optimal length for a YES/NO vocabulary test?

Hypothesis 3

Scores on a 120-real-word checklist vocabulary test (Test C) will correlate more strongly and more consistently

with CELT subtest scores than will scores on two 60-real-word checklist vocabulary tests (Test A and Test B) for both vocabulary test scoring procedures--the simple-checklist procedure and the YES/NO procedure.

This hypothesis explores the interaction between sample size and test procedure. As sample rate increases the content validity of the test increases, but as the test becomes longer there is a potential for decreases in subject performance due to fatigue or boredom.

Research Question 3

Can the level of L2 proficiency needed by EAP students in order to exit EAP programs be partially described in terms of necessary knowledge of the most-frequent 5,000 word families of English?

Hypothesis 4

Subjects who score above recommended EAP program exit scores on the CELT subtests (CELT Listening = .80; CELT Listening/Structure avg. = .73) will also score above 5,000 on the checklist vocabulary tests.

This hypothesis is based on research that suggests that knowledge of the most-frequent 5,000 word families of English is necessary in order to comprehend most English text.

CHAPTER II

REVIEW OF LITERATURE

L2 ABILITIES NEEDED BY EAP STUDENTS

The academic success of international students in the U.S. depends on much more than their English language abilities. For instance, a student's financial situation, family situation, or personal health situation can all affect his or her academic success in dramatic ways. However, language ability is a critical factor in academic success, and it is the primary goal and responsibility of EAP programs to help their students acquire the type and level of second language ability that they will need for future academic success in English-medium institutions of higher education.

There is some consensus that, of the 'four-skills' abilities (reading, writing, listening, speaking), reading, writing, and listening are the most crucial L2 needs of EAP students (Sloan, 1994). There is also some consensus concerning the L2 needs of EAP students with respect to underlying knowledges or competencies. In this case there is wide agreement that lexical knowledge is a very important need of EAP students and some consensus that lexical knowledge may be the greatest need of EAP students (Gass & Selinker, 1994). For example, Goulden, Nation, & Read (1990)

state that "measures of vocabulary size . . . are important indicators of the ability of second language learners to achieve academic success" (p. 342). Saville-Troike (1984) claims that "vocabulary knowledge is the single most important area of second language (L2) competence when learning content through that language is the dependent variable" (p. 199).

Thus, research on the relationship between EAP students' lexical knowledge at various levels of L2 abilities may be helpful in describing the SLA process for EAP students. The research questions presented above are thus important. Some interesting research has been done recently that addresses these questions. This thesis hopes to add to this body of knowledge.

DEFINING AND COUNTING WORDS

Studies of vocabulary size, the number of words in texts, and word frequency have been published for a number of years. In most of these studies, researchers have counted or ranked graphically distinct words--'graphic words.' A 'graphic word' is "a string of alphanumeric characters with space on either side; may include hyphens and apostrophes but no other punctuation marks" (Francis & Kucera, 1982: 3). Some recent research suggests that grouping related graphic words into 'word families' may better reflect the organization of the mental lexicon and may be a useful

construct in vocabulary research (Carver, 1994; Nation, 1993; Read, 1988). Recent L1 research done with children shows that word-family frequency can be a significantly better predictor of word knowledge than graphic-word frequency (Ryder & Slater, 1988), and recent counts of the number of words in texts and estimates of adult, native-speaker vocabulary size based on the concept of 'word family' are more consistent with each other than are previous studies based on graphic words and therefore appear more valid (D'Anna, Zechmeister, & Hall, 1991; Goulden, Nation, & Read, 1990).

As a construct, a 'word family' is a group of graphic words that are 'semantically transparent.' In other words, it is a group of graphic words such that knowledge of any specific graphic word in the family will help in understanding any other graphic word in the family. Graves et al. (1987) present the following group of graphic words as an example of a word family:

beautifully	beauties	beautified
beautiful	Beautiful	beautify
beauty	Beauty	beauty's
Beauty's	beaut	beauteous
beautification	beautifulest	beautifying

Currently, operational definitions of the construct vary slightly from researcher to researcher. However, in the following review, studies based on the concept of word family are assumed to be comparable.

L2 VOCABULARY KNOWLEDGE AND EAP ABILITY

Reading

It is widely acknowledged that vocabulary knowledge is an important, even critical feature of L2 reading ability (Grabe, 1991). Word difficulty is the most important factor in all measures of readability (Anderson & Freebody, 1983). However, some researchers have claimed that the single greatest need of most adult L2 readers is a larger vocabulary (Fox, 1979; Kelly, 1990; Laufer, 1992; Laufer & Sim, 1985; Meara, 1982; Strother & Ulijn, 1987; Ulijn, 1984). This statement has also been made by L2 readers themselves. For example, Gorman's (1979) survey of advanced EAP students at UCLA revealed that two-thirds of the students polled believed that inadequate vocabulary size was their biggest problem with academic reading.

If vocabulary size is so important for reading, how much is enough? This question has been asked and answered in two ways. First, researchers have asked what percentage of the words in a text a reader needs to 'know' in order to comprehend the text adequately. Second, researchers have asked how many word families it takes to provide various levels of coverage of various types of English text.

There is currently some consensus that if fewer than 95% of the word families in a written text are known understanding will be based much more on guessing than on

confident comprehension (Kyongho & Nation, 1989; Laufer, 1989; Liu & Nation, 1985; Ostyn & Godin, 1985). However, other research claims that the percentage of word families that need to be known is even higher--that 98% of the words in a text need to be known for normal reading (Bright & McGregor, 1970; Carver, 1994; Hirsh & Nation, 1992; Nation, 1994).

There is also some consensus that it takes the most-frequent 5,000 word families of English to achieve 95-98% coverage of the words in most written English texts (Fox, 1979; Hirsh & Nation, 1992; Kyongho & Nation, 1989; Laufer, 1989, 1992; Ostyn & Godin, 1985). Thus, there seems to be some agreement that a 5,000-word-family vocabulary is needed to read most English texts 'normally'--with confident comprehension and without frustration. Laufer (1989, 1992) explains this 5,000-word-family vocabulary as a threshold below which top-down reading processes will not be effective. Other researchers posit and explain 95-98% lexical coverage as a threshold below which it is very difficult to guess the meanings of unknown words from context (Liu & Nation, 1985) or below which reading is much more likely to be frustrating rather than pleasurable (Carver, 1994; Hirsh & Nation, 1992).

The research findings presented above provide some preliminary answers to the first two questions that this thesis addresses (see page 6.) The findings suggest that the relationship between lexical knowledge and reading ability

may not be strictly linear in the sense that greater lexical knowledge equals greater reading ability. Rather, it suggests that there may be a lexical threshold above which other knowledge necessary to the reading process will become available and reading ability will increase not linearly but exponentially. If this hypothetical 'lexical threshold' for reading ability can be described in terms of knowledge of the 5,000 most-frequent word families in English, then all but the most advanced EAP students must have not yet acquired this knowledge. Perhaps they are reaching this level of lexical knowledge just around the time in their English development that they exit EAP programs. This hypothesis will be tested in this thesis.

Writing

Although there is little research on the relationship between the L2 vocabulary knowledge and the L2 writing ability of adult EAP students, the results of two recent studies suggest an important relationship.

Ferris (1991, 1994) analyzed a corpus of essays written by 160 EAP students at a university in the U.S. The essays were written as part of a placement-test battery and were holistically scored on a scale of 1 to 10 by three independent teacher-raters. Ferris found that of 62 quantitative syntactic and lexical text variables studied,

'number of word tokens' in the text was the best predictor of holistically determined essay scores.

Leki and Carson (1994) surveyed 77 international students who were enrolled in their first year of non-ESL university classes. All of the subjects had completed EAP programs the previous year. In an open-ended question, the subjects were asked what they wish they had learned or learned better in their previous EAP writing classes in order to help them with the writing required in their current non-ESL university classes. Leki and Carson found that "respondents mentioned vocabulary more frequently than they did any other single feature of their writing needs" (p. 90). Leki and Carson present the following quote from a student response as representative:

I wish I had learned more words and had increased my vocabulary significantly. Sometimes I simply run out of words necessary to express what I am actually thinking. I usually find a substitute word, but often times that word would only be semi-fitting with my thoughts (1994: 91).

Listening

Other research has investigated the relationship between lexical knowledge and L2 listening ability. Nation (1994) claims that the most-frequent 2,000 word families of English provide around 95% lexical coverage of spoken English. If this is so, then a 2,000-word-family vocabulary should be sufficient to understand much of spoken English. However, not all spoken English is the same, and it may be that the

vocabulary size needed to understand lecture-type spoken English, an important ability for EAP students, is much greater than that needed to understand casual face-to-face conversation.

Kelly (1991) analyzed transcriptions of recorded BBC radio newscasts made by 38 advanced, francophone EFL students. After identifying and counting different types of errors, Kelly concluded that "[t]he data on foreign language learner misperceptions . . . point conclusively to lexical ignorance as the main obstacle to listening comprehension with advanced foreign language learners" (p. 147). Some evidence of why this may be so is provided by Hayes and Ahrens' (1988) study of a corpus of various types of spoken texts. Table I shows the lexical coverage provided by the most-frequent 3,000 word families in English for various types of texts in Hayes and Ahrens' corpus. Hayes and Ahrens actually calculated the lexical coverage provided by the 5,000 most-frequent graphic words. It is assumed here that the 5,000 most-frequent graphic words of English consists of approximately 3,000 word families. This estimate is based on Goulden's (1990) conclusion that the 10,000 most-frequent graphic words in Thorndike and Lorge's (1944) word-frequency list consists of 6,100 word families.

TABLE I
 LEXICAL COVERAGE PROVIDED BY THE APPROXIMATELY
 3,000 MOST-FREQUENT WORD FAMILIES
 OF ENGLISH FOR VARIOUS TYPES OF ORAL TEXT

	<u>% of text covered</u>
I. TELEVISION TEXTS	
a. Popular prime-time adult shows	94
b. Popular prime-time children's shows	93
c. Cartoon shows	92
d. Mr. Rogers and Sesame Street	97
II. ADULT SPEECH	
a. Expert witness testimony	90
b. College graduates to friends, spouses	94

Note. Adapted from Hayes and Ahrens (1988).

There are several interesting findings shown in Table I. An approximately 3,000-word-family vocabulary does not provide greater than 95% coverage of any of the oral texts except for Mr. Rogers and Sesame Street programs. A 3,000-word-family vocabulary is close to providing greater than 95% coverage of the face-to-face conversational English texts. However, a 3,000-word vocabulary is far from providing adequate coverage of the more lecture-type adult speech that was analyzed ('expert witness testimony'). The question raised by these results is: What number of word-families would it take to achieve greater than 95% lexical coverage of the more lecture-type 'expert witness testimony' texts? Perhaps it would take a number similar to that needed for greater than 95% coverage of most written texts--5,000.

The research presented above concerning the relationship between lexical knowledge and L2 listening ability is consistent with the research on the relationship between lexical knowledge and L2 reading ability. It suggests that lexical knowledge is a very important need for EAP students. It also suggests that the L2 listening ability needed by these students--the ability to comprehend unsimplified university level lectures in English--may require a 5,000-word-family vocabulary. Helping EAP students acquire the ability to understand unsimplified lectures in English is an important goal for most EAP programs, and students do generally acquire this ability close to the time that they exit EAP programs. Perhaps attainment of the L2 listening ability needed to exit EAP programs closely coincides with the attainment of knowledge of the 5,000 most-frequent word families of English. This hypothesis will be tested in this thesis.

Testing the above hypotheses requires the measurement of students' lexical knowledge. How can we measure the lexical knowledge of EAP students?

VOCABULARY TESTS

Many useful tests exist for testing EAP students' lexical knowledge. Many of these tests are subcomponents of standardized proficiency tests such as the Comprehensive English Language Test (CELT) or the TOEFL. Other types of

vocabulary tests also exist such as Nation's (1990) 'vocabulary levels test,' a very useful diagnostic test intended to give teachers a good general idea of the number of words known by their students. However, none of these tests seems to be well-suited to testing the hypothesis that the attainment of knowledge of the 5,000 most-frequent word families of English closely coincides with the attainment of the L2 abilities required to exit EAP programs.

An ideal vocabulary test to be used in testing this hypothesis would be created by first choosing a word-family-frequency list. A word-family-frequency list is created by first counting the frequency of graphic words in a large corpus. Second, graphic words are grouped into families. Third, the frequency of each word family is calculated by summing the frequencies of each of its graphic-word members. Fourth, the word families are rank ordered. Finally, each word family is represented in a rank-ordered list by its base word.

With such a list, each word family can be represented in a vocabulary test by its base word. Test words should be chosen by taking a systematic, stratified sample from the first five or six thousand words in the rank-ordered list. This can be done by choosing a random starting point in the first few words of the test and then choosing every 'n'th word throughout the list.

Table II shows that in statistical terms the larger the sample size the better. Table II shows the predicted 95% confidence intervals for an individual subject's score on various size hypothetical tests sampled from a 5,000-word source list if the subject obtained a score of 50% (Fountain, 1995). This is the most conservative prediction possible. The actual confidence interval depends on the subject's score. As a subject's score approaches either zero or one, the confidence interval narrows.

TABLE II

95% CONFIDENCE INTERVALS FOR A 50% SCORE ON VARIOUS
SIZE VOCABULARY TESTS SAMPLED FROM A 5,000-WORD LIST

# of test words	95% Confidence Interval in % of source list	95% Confidence Interval in # of source-list words
60	.50 +/- .13	2500 +/- 650
120	.50 +/- .09	2500 +/- 450
150	.50 +/- .08	2500 +/- 400
200	.50 +/- .08	2500 +/- 400
250	.50 +/- .06	2500 +/- 300
300	.50 +/- .05	2500 +/- 250
350	.50 +/- .05	2500 +/- 250
450	.50 +/- .04	2500 +/- 200
600	.50 +/- .04	2500 +/- 200
1200	.50 +/- .02	2500 +/- 100

Table II shows, for example, that if a subject obtained a score of 50% on a 250-word vocabulary test where the test words were taken from a 5,000-word source list, there is 95% certainty that the true number of words in the 5,000-word source list that the subject knows is somewhere between 2,200 and 2,800. Table II shows that testing the number of words known in a 5,000-word source list requires a large sample in order to state with high levels of confidence the true number of words known by the subject.

Table II is based of the assumption that the procedure for testing knowledge of individual test words is perfect. This is not the case. Moreover, sample size and test procedure interact and place complicated practical constraints on sample size. For example, the multiple-choice procedure for testing word knowledge requires a large amount of time per word to complete. Multiple-choice tests with only 75 items take around 35 minutes for EAP students to complete (Harris & Palmer, 1986). The checklist procedure requires much less time per word. However, it does not require subjects to prove their knowledge of a word which may invite questions as to its validity. Moreover, it seems reasonable to suspect that as the length of a checklist test increases, the validity of the procedure decreases. As subjects become physically tired during, or bored with, the test, their level of cautiousness in answering items may decrease. Finally, if the test is too long, the researcher may not be granted time to give the test or may not have enough time to use multiple tests.

A relatively new type of testing procedure known as the 'YES/NO' procedure may be a useful way to test a large number of words in a short amount of time. YES/NO tests contain both real English words and nonsense words. The percentage of real words marked and the percentage of nonsense words marked are applied to a mathematical formula to determine a

subject's test score. With this procedure 150 words can be tested in less than ten minutes.

YES/NO VOCABULARY TESTS

Overview of the YES/NO test procedure

Beginning with the work of Zimmerman et al. (1977), this testing procedure has been evolving over the last 25 years. The main hypothetical strength of this procedure is that, like simple-checklist tests and unlike multiple-choice tests, it takes a short amount of time to answer an item. Thus, a large number of words can be tested compared to multiple-choice tests. This provides the potential to create vocabulary tests that are more sensitive than current multiple-choice tests. A second hypothetical strength of this procedure is that it also accounts for the potential for subjects, on simple-checklist tests, to mark words that they do not really know.

Below is an example of what a YES/NO test looks like. This is only an example and is not part of a real test. A real YES/NO vocabulary test would need to be much longer and the directions would need to be different for low-level ESL students.

Example YES/NO Vocabulary Test.

Directions: The following list contains some common English words, some uncommon English words, and some 'nonsense' words. Mark the words that you know.

stipeld	mirror	convict
preventive	acknowledge	bee
listure	outside	sustend
suspect	vast	resistance
candigent	schedule	excind

The above example YES/NO test contains 15 words. Two-thirds of the words are real English words that can be found in the Oxford American Dictionary. Most ESL/EFL students would not mark all of the real words and would mark at least one nonsense word. Marking a nonsense word can be, but is not necessarily an indication of 'lying.'

The combination of REAL and NONSENSE words and YES (checked) and NO (not checked) answers yields four possible response outcomes:

- (1) The subject can respond YES to a REAL word--a YES/REAL response. (This is also known as a 'HIT.')

(2) The subject can respond NO to a REAL word--a NO/REAL response.

(3) The subject can respond YES to a NONSENSE word--a YES/NONSENSE response. (This is also known as a 'False Alarm'.)

(4) The subject can respond NO to a NONSENSE word-- a NO/NONSENSE response.

Although there are four possible responses, in scoring the test only two of these possibilities are important--the YES/REAL responses (the 'Hits') and the YES/NONSENSE responses (the 'False Alarms').

Before the procedure used to score a YES/NO test is explained, some examples of possible performances on a YES/NO test will be presented.

Example YES/NO test performance #1 -- Subject A.

Suppose Subject A answers YES to all of the real words on the test. This is a Hit Rate of 100%. Suppose Subject A also answers NO to all of the nonsense words. This is a False Alarm Rate of 0%. These results tend to indicate that Subject A is able to distinguish perfectly between all of the real and nonsense words. Results like this indicate that the test was too easy.

Example YES/NO test performance #2 -- Subject B.

Suppose Subject B has a Hit Rate of 100% and a False Alarm Rate of 100%. Subject B then has not demonstrated any ability to distinguish between the real and the nonsense

words. There is no evidence therefore that the subject knows any of the real words.

Example YES/NO test performance #3 -- Subject C.

Suppose Subject C has a Hit Rate of 50% and a False Alarm Rate of 0%. Subject C then claims to know some of the real words and none of the nonsense words. This type of score is assumed to be a fairly accurate self-assessment.

Example YES/NO test performance #4 -- Subject D.

Suppose Subject D, like Subject C, claims to know half of the real words--a Hit Rate of 50%. However, suppose Subject D also claims to know 10% of the nonsense words--a False Alarm Rate of 10%. This indicates that Subject D's criteria for knowing a word may be somewhat lenient and that a score slightly lower than 50% might more accurately reflect Subject D's knowledge of the test words.

Performances such as Subjects D's are by far the most common. Thus, nearly all subjects' Hit Rates will, according to the theory, need to be adjusted to obtain a more accurate estimate of the real number of words known.

There have been a few cases reported where a subject's False Alarm Rate was higher than his or her Hit Rate. Meara, Lightbown, & Halter (1994) suggest that this may be due to the subject's guessing wildly, perhaps as a result of a test that is too difficult.

A mathematical formula is used to adjust a subject's score. The adjustment formula (translated into the terminology used here) is as follows:

$$PK = \frac{\text{Hit Rate} - \text{False Alarm Rate}}{1 - \text{False Alarm Rate}}$$

The Hit Rate is the number of 'Hits' (i.e., YES/REAL responses) divided by the number of real words in the sample. The 'False Alarm Rate' is the number of 'False Alarms' (i.e., YES/NONSENSE responses) divided by the number of nonsense words in the test. 'PK' ('percentage known') is the estimated percentage of real words in the test known by the subject. 'PK' can be multiplied by the number of words in the sampled source list to give an estimate of the total number of words in the sampled list known by the subject.

A key point of the theory is that, within a certain range, it is a subject's knowledge of the real words, not the subject's level of cautiousness, that determines the subject's PK score. As a subject becomes less cautious, that subject's Hit Rate should increase, but that subject's False Alarm Rate should also increase. If the subject becomes more cautious, just the opposite should happen. Either way, the subject's PK score should remain about the same.

Development, use, and validation of 'YES/NO' vocabulary tests

L1 Studies. Checklist vocabulary tests that include some nonsense words were first developed and studied by Zimmerman

et al. (1977) with American, university-student subjects. Zimmerman et al. concluded that such a method was very promising.

Building on the Zimmerman et al. research, Anderson and Freebody (1983) were the first to create and examine the validity of a true YES/NO test. Using a YES/NO test containing 195 real words and 131 nonsense words, Anderson and Freebody studied fifth-grade, native-speaker-of-English students. Their research yielded two interesting findings. First, there was a strong correlation (.84) between scores on the YES/NO test and scores on a multiple-choice test of the same target words. Second, and more importantly, Anderson and Freebody found that correlation with an oral interview test of the same target words was much higher for the YES/NO test (.85-.92) than for the multiple-choice test (.43-.45).

Cooksey and Freebody (1987) replicated the Anderson and Freebody study with a few modifications. They found more evidence for the reliability and validity of the YES/NO test method. Cooksey and Freebody studied 4th-, 7th-, and 10th-graders--43 subjects in all. All of the subjects were native speakers of English. The students were given a computerized YES/NO test that contained 220 real words and 120 nonsense words.

Evidence of YES/NO test validity was sought by comparing the YES/NO test scores with the results of an interview vocabulary test of the same target words and with the known,

grade and general-ability level of the students. Cooksey and Freebody concluded that the YES/NO test had a "moderate degree of convergent validity with the interview procedure and a high degree of known groups validity" (p. 116).

White, Slater, and Graves (1989) studied 278 elementary-school children. The subjects were all native speakers of English. White et al. compared the results of a YES/NO test, a multiple-choice test, and an interview test of the same target words. The YES/NO test contained 56 real words and 44 nonsense words. The YES/NO test was a better predictor of performance on the oral interview of test-word knowledge than was the multiple-choice test. White et al. (1989) concluded that "yes/no testing offers a highly efficient and accurate way to estimate the average proportion of words that students can read and understand" (p. 397).

L2 Studies. The results of L2 research on YES/NO tests resemble the results of research done in L1 contexts. Although few studies have been done, all of the studies suggest promise for the YES/NO method.

Meara and Buxton (1987) were the first to publish a study of the use of a YES/NO vocabulary test with ESL/EFL students. Meara and Buxton gave 100 adult ESL students a 25-item multiple-choice test similar to the vocabulary test used in the Cambridge First Certificate Examination. They also gave their subjects a YES/NO test that consisted of 60 real words and 40 nonsense words. Correlation between scores on

the two tests was quite high even though the target words were different in each of the tests. Twenty-six of the subjects also took the Cambridge First Certificate in English examination. Meara and Buxton compared their subjects' scores on the Cambridge test with their subjects' scores on the multiple-choice test and their subjects' scores on the YES/NO test. They concluded that the YES/NO test was a much better predictor of Cambridge test scores than the multiple-choice test.

In a follow-up study, Meara and Jones (1988) created a computerized YES/NO test and tested it as a potential in-house placement test for the 'Eurocentres' chain of language schools. The YES/NO test consisted of a stratified sample taken from the first 10,000 words of Thorndike and Lorge's (1944) graphic-word-frequency list (Read, 1988). The first 10,000 words of Thorndike and Lorge's list consists of around 6,000 word families (Goulden et al., 1990). The YES/NO test was given to 250 adult ESL students who also took the Eurocentres' in-house placement test--the Joint Entrance Test (JET). The JET, which was designed to test for the level and type of L2 proficiency needed by EAP students, is similar to the Michigan test, the CELT, the TOEFL, and other such tests and consists of a listening comprehension subtest, a grammar subtest, a reading subtest, and an oral interview.

Correlation between YES/NO test scores and JET scores ranged from .55 - .80 for different L1 groups. The YES/NO

test was a better predictor of the subjects' class levels than the JET based on initial placements of students into Eurocentres' classes. The YES/NO test was an even better placement indicator when subsequent adjustments to initial placements were taken into account. The fact that correlation between the YES/NO test and the JET varied for different L1 groups raised some questions. Meara hypothesized that this variation might be related to L1-L2 cognates in the YES/NO test.

Meara, Lightbown, and Halter (1994) tested the effect of L1-L2 cognates on the outcome of a YES/NO vocabulary test. They tested 89 adult ESL students who were native speakers of French. Meara et al. gave the subjects two different YES/NO tests, the CELT Listening subtest, and the CELT Structure subtest. Both YES/NO tests contained 60 items--40 real words and 20 nonsense words. The real words were taken from the 5,000 most-frequent words in Kucera and Francis' (1967) graphic-word-frequency list. Fifty percent of the real words in one of the YES/NO tests were cognate with French. The other YES/NO test did not contain any words that are cognate with French. The difference between the mean score on the YES/NO test that contained French cognates (COG Test) and the YES/NO test that did not contain any cognates (NCOG Test) was statistically significant, but not large. It amounted to a difference of about 300 words in the total vocabulary size estimate. Meara et al. combined the two YES/NO tests for

each subject and analyzed the combination as if it were a twice-as-long third test--the cognate plus noncognate YES/NO test (COG + NCOG Test). They calculated the correlation between their subjects' scores on the three YES/NO tests and their subjects' scores on the two CELT subtests.

Table III shows that all three YES/NO tests were fairly good predictors of CELT scores.

TABLE III
CORRELATION BETWEEN YES/NO TESTS AND
CELT SUBTEST SCORES FROM MEARA ET AL. (1994)

<u>YES/NO Tests</u>	<u>CELT Subtest Scores</u>	
	Listening	Structure
COG Test	.66	.66
NCOG Test	.74	.76
COG + NCOG Test	.75	.76

Note. From Meara et al. (1994).

Table III shows that the COG + NCOG test did not correlate more strongly with the CELT subtests than the NCOG test did independently. This is interesting since it contained twice as many items. This may suggest that the point of diminishing returns for test-word sample size is quite low. However, the negative effect of cognates may have simply canceled out any positive effect of increased sample rate.

The validation studies of the YES/NO test procedure reviewed above consistently concluded that YES/NO tests can

be better measures of vocabulary size than multiple-choice tests. However, none of these studies tested whether the YES/NO procedure was better than the simple checklist procedure. Such a study might be useful.

RATIONALE FOR THE PRESENT STUDY

It has been shown through the literature reviewed above that the development of and use of new types of vocabulary tests might be useful research. A useful test might test a large, stratified sample of words taken from a rank-ordered word-family frequency list using both the simple-checklist procedure and the YES/NO procedure. Use of such a test might help in answering the following research questions:

1. Is the YES/NO test procedure a more valid measure of L2 vocabulary knowledge than the simple-checklist procedure?
2. What is the optimal length for a YES/NO vocabulary test?
3. Can the level of L2 proficiency needed by EAP students in order to exit EAP programs be partially described in terms of necessary knowledge of the most-frequent 5,000 word families of English?

This study attempts to create and examine the efficacy of a simple-checklist and YES/NO vocabulary tests and in the

process provide some empirical data that will help in answering the above research questions.

CHAPTER III

METHOD

SUBJECTS

The subjects for this study were 66 students enrolled in an EAP program at a four-year college in the United States. The group of subjects was heterogeneous with respect to L1. The breakdown according to L1 was as follows: Japanese (N=32), Arabic (N=23), Korean (N=7), Indonesian (N=3), Chinese (N=1). Some of the students had been studying English in the U.S. for at least a year while others had been in the U.S. for only a few months.

For various unavoidable and largely unknown reasons (perhaps due to absence from class on the day the test was given or perhaps due to preferring not to participate), not all of the students in the EAP program participated in this study. However, completed vocabulary tests and CELT Listening and Structure subtest scores were obtained for approximately 75% of the students in the program. Since the subjects' L2 proficiency ranged from 'high-beginner' to 'advanced,' and since the nonparticipation of some students in the study did not appear to be systematic, it is assumed that the group 66 students that participated in this study were representative of the EAP program as a whole.

INSTRUMENTS

Two parallel vocabulary tests were created as follows. First, the adjusted-rank lemma-frequency list from Francis and Kucera (1982) was chosen as the list to be sampled. Thus, the operational definition of 'word family' chosen for this study was the definition of a 'lemma.' A lemma is a group of related graphic words that consists of a base word and its inflectional variants that belong to the same part of speech. In this list, the 6,000 most-frequent lemmas in written American English are represented by their base word. The base words are listed in descending order of frequency. The list is based on Kucera and Francis' (1967) word count of a 1,014,000-graphic-word corpus of written American English.

Two 90-item tests were created. Each test consisted of 60 words sampled from Francis and Kucera's (1982) 'adjusted-rank' list and 30 nonsense words created by the researcher. For Test A, a stratified, random sample of real words was created by choosing every 100th word from the adjusted-rank list beginning with the 100th most frequent word--a total of 60 words ($6,000/100 = 60$). A word was rejected if it had a more frequently occurring homograph. This was necessary as YES/NO tests depend on visual recognition of decontextualized words and homographs often have very different frequencies (e.g., 'book' as a noun and 'book' as a verb). If a word was rejected because it had in the list a more frequently

occurring homograph, the next, more-frequent word that met the criterion was chosen.

For Test B, the same sampling procedure used for Test A was followed except that the sample began with the 99th rather than the 100th most-frequent word. Test A and Test B do not contain any of the same words, and the frequency distributions of the test words in Test A and Test B were as identical as possible.

The final samples of real words for both Test A and Test B were analyzed to see how many words were one-syllable words, how many words were two-syllable words, and so on. The nonsense words were constructed so that the ratio of one-syllable nonsense words to total nonsense words for each test would be the same as the ratio of one-syllable real words to total real words. The same was done for two-syllable words, three-syllable words, and so on.

After a list of nonsense words was created, opinion was solicited from L1 speakers of English as to which of the nonsense words seemed most like real English words and to make sure that no meaning could be attributed to them. New words are coined every day, and the meaning of some of these words is immediately understandable at first sight, even without context. Nonsense words of this type were carefully avoided. The final group of nonsense words was checked to make sure that none of the words could be read as homophones of real English words. For example, the nonsense word

'baire' could not be used since it can easily be read as a homophone of the real English words 'bear' and 'bare.' Finally, the Oxford American Dictionary was consulted to confirm that the nonsense words were in fact not real English words.

Nonsense words were randomly assigned to either Test A or Test B. This was done by writing each nonsense word on a 3 x 5 note card. The cards were then sorted into five piles--a pile for one-syllable nonsense words, a pile for two-syllable nonsense words, and so on. The piles were shuffled, and the appropriate number of words was dealt to each test.

The real and nonsense words for each test were then randomly listed. This was done for each test by simply writing each real word and each nonsense word on a 3 x 5 note card, shuffling the cards, and then numbering and listing the words in the order of the cards. The words in Test A were numbered 1-90. The words in Test B were numbered 91-180. Test A appears as Appendix A. Test B appears as Appendix B. An analysis of Test A and Test B showing the rank of each real test word in Francis and Kucera's list appears as Appendix C.

PROCEDURES

The tests were administered to intact groups of students in class by their teachers, who at a faculty meeting had been

given a short briefing on the project and instructions on how to give the test. The two YES/NO tests were given together so that they appeared to be one 180-word test. Pilot studies suggested that it would take about seven minutes for the subjects to complete both 90-word tests and that the total testing procedure would take around 15 minutes, including information, instructions, consent forms, and the test itself.

The subjects were told that the test contained some real English words and some words that look like English words but are not. The subjects were not told what the proportion of real to nonsense words was. The subjects were asked simply to circle the words that they 'know' the meaning of. The directions were presented in written English at the beginning of the test. The test directions are shown in Appendix D.

All of the subjects had taken the 'Structure' and 'Listening' subtests of the Comprehensive English Language Test (CELT) several weeks prior to taking the YES/NO tests. Scores on these CELT subtests were obtained for all 66 subjects. The CELT has been shown to be a valid general measure of the L2 ability of intermediate and advanced ESL/EFL students in high school, college, and adult programs (Harris & Palmer, 1986). It is widely used as a placement test in EAP programs in the U.S.

HYPOTHESES AND ANALYSES

Test scores

Test A and Test B were scored using both the YES/NO procedure and also as simple-checklist tests. Test A and Test B were also be combined and scored as one test, creating a third vocabulary test--Test C. Therefore, there were six vocabulary test scores--three tests times two scoring procedures. There were three 'CELT' scores. CELT Listening and Structure subtest scores were averaged creating a third CELT subtest score--'CELT -L/S.' The complete CELT contains three sections. The third section is a vocabulary subtest that was, unfortunately, unable to be administered. Thus, the creation of CELT-L/S scores was an attempt to approximate total CELT scores.

Hypothesis 1

Scores on Test A, Test B, and Test C will correlate more strongly and more consistently with CELT Structure subtest scores and CELT Listening subtest scores when the vocabulary tests are scored using the YES/NO procedure than when scored as simple checklist tests.

Test of Hypothesis 1

This hypothesis was tested by calculating the correlation among all three CELT subtest scores and all six vocabulary test scores for all 66 subjects.

Hypothesis 2

The correlation among checklist vocabulary test scores and CELT subtest scores for subjects with 'high' False Alarm Rates (greater than 10%) versus subjects with 'low' False Alarm Rates (less than or equal to 10%) will be meaningfully different for simple-checklist scores but not for YES/NO scores.

Test of Hypothesis 2

This hypothesis was tested by dividing the subjects into two groups--a group with 'high' False Alarm Rates (greater than 10%) and a group with 'low' False Alarm Rates (less than or equal to 10%). Correlation among vocabulary test scores and CELT subtest scores was examined. The theory behind the YES/NO procedure claims that different levels of subject cautiousness will affect simple-checklist scores to a much greater degree than they will affect YES/NO test scores.

Hypothesis 3

Scores on Test C will correlate more strongly and more consistently with CELT subtest scores than will scores on Test A and Test B for both vocabulary test scoring procedures--the simple checklist procedure and the YES/NO procedure.

Test of Hypothesis 3

This hypothesis was tested by calculating the correlation among all three CELT subtest scores and all six vocabulary test scores for all 66 subjects.

Hypothesis 4

Subjects who score above recommended EAP program exit scores on CELT subtests (CEL_T Listening = .80; CEL_T-L/S = .73) will also score above 5,000 on the checklist vocabulary tests.

Test of Hypothesis 4

This hypothesis was tested by plotting the correlation among Test C scores, CEL_T-L/S scores, and CEL_T Listening scores. It is assumed that the CEL_T - L/S scores give a fairly good estimate of what a subject's score would be if the subject had taken all three subsections of the CEL_T. According to The Guide: A Resource for International Admission Professionals (1994), "in general, [CEL_T] scores above 220 are considered satisfactory, whereas scores below 180 . . . indicate the need for fulltime instruction" (p. 105). As CEL_T subtest scores represent the percentage of possible correct answers on that section, total CEL_T scores can also be represented as percentages as was done in this study. A score of 220 can therefore be represented as 220/300 or .73, and a total score of 180 can be represented as 180/300 or .60. The Guide (1994) also suggests that "[i]f

only the listening portion [of the CELT] is used[,] . . . [a] score of 80 or higher indicates the test taker may be able to successfully process academic lectures in certain fields, particularly those in mathematics, computer science, and engineering" (p. 105).

CHAPTER IV

RESULTS

Research Question 1

Is the YES/NO test procedure a more valid measure of L2 vocabulary knowledge than the simple-checklist procedure?

Hypothesis 1

Scores on Test A, Test B, and Test C will correlate more strongly and more consistently with CELT subtest scores when the tests are scored using the YES/NO procedure than when scored as simple checklist tests.

Table IV shows the correlation among the CELT subtest scores and the vocabulary-test scores for all 66 subjects. (The complete correlation matrix is presented as Appendix E.)

TABLE IV

CORRELATION AMONG ALL THREE CELT SUBTESTS AND
ALL THREE VOCABULARY TESTS FOR ALL SUBJECTS (N=66)

	<u>CELT-S</u>	<u>CELT-L</u>	<u>CELT-L/S</u>
Test A-SC	.60	.57	.64
Test B-SC	.60	.61	.66
Test C-SC	.62	.61	.67

Test A-YN	.44	.39	.46
Test B-YN	.63	.62	.71
Test C-YN	.63	.56	.65

Key:

CELT-S=CELT Structure subtest scores

CELT-L=CELT Listening subtest scores

CELT-L/S=scores created by averaging a subject's CELT-S and CELT-L scores

Test A-SC=the percentage of real words circled on Test A (i.e., simple checklist score)

Test B-SC=the percentage of real words circled on Test B (i.e., simple checklist score)

Test C-SC=the percentage of real words circled on Test C (i.e., simple checklist score)

Test A-YN=Test A scored using the YES/NO procedure

Test B-YN=Test B scored using the YES/NO procedure

Test C-YN=Test C scored using the YES/NO procedure

The top three rows of Table IV show the correlation among all three vocabulary tests scored as simple-checklist tests and all three CELT subtest scores.

There was no meaningful difference in the correlation among all three vocabulary tests scored as simple-checklist

tests and CELT Structure subtest scores. The correlation ranged from .60 to .62 (column 1, rows 1, 2, 3).

There was also no meaningful difference in the correlation among all three vocabulary tests scored as simple-checklist tests and CELT Listening subtest scores. The correlation ranged from .57 to .61 (column 2, rows 1, 2, 3).

There was also no meaningful difference in the correlation among all three vocabulary tests scored as simple-checklist tests and CELT-L/S scores. The correlation ranged from .64 to .67 (column 3, rows 1, 2, 3).

Thus, the correlation among CELT subtest scores and all three vocabulary tests scored as simple-checklist tests does not seem to be affected by which vocabulary test is used. There is a meaningful difference according to CELT subtest though. The correlation among the vocabulary tests scored as simple-checklist tests and the CELT-L/S scores (.64 to .67) (column 3, rows 1, 2, 3) is consistently stronger than the correlation among the vocabulary tests scored as simple-checklist tests and the CELT Listening subtest and CELT Structure subtest scores (.57 to .62) (columns 1 and 2, rows 1, 2, 3).

Rows 4, 5, and 6 show the correlation among the three vocabulary tests scored as YES/NO tests and the CELT subtests.

There was a meaningful difference in the correlation among the three vocabulary tests scored as YES/NO tests and

among the three vocabulary tests scored as YES/NO tests and CELT Structure subtest scores. The correlation ranged from .44 to .63 (column 1, rows 4, 5, 6).

There was also a meaningful difference in the correlation among the three vocabulary tests scored as YES/NO tests and the CELT Listening subtests. The correlation ranged from .39 to .62 (column 2, rows 4, 5, 6).

There was also a meaningful difference in the correlation among the three vocabulary tests scored as YES/NO tests and CELT-L/S scores. The correlation ranged from .46 to .71 (column 3, rows 4, 5, 6).

The patterns in Table IV can be summarized by saying that the correlation among the CELT subtests and the vocabulary tests is more consistent and stronger when the vocabulary tests are scored using the simple-checklist procedure than when they are scored using the YES/NO procedure.

Meara (1992), who has used YES/NO vocabulary tests extensively, claims that if a subject's False Alarm Rate is higher than 50%, the score-adjustment procedure will not work properly. There were five subjects who had unusually high False Alarm Rates (#s 39, 41, 46, 57, 59). These subjects had False Alarm Rates ranging from .42 to .87 on Test C. (Data for all subjects is shown as Appendix F.) One of these subjects (#41) had a False Alarm Rate that was greater than his or her Hit Rate on Test A and thus a YES/NO score of zero. Table V shows the correlation among the CELT subtests

and the vocabulary tests after these five anomalous scores were edited from the data. (The complete correlation matrix is shown as Appendix G.)

TABLE V
CORRELATION AMONG CELT SUBTESTS AND
VOCABULARY TESTS FOR ALL SUBJECTS EXCEPT FIVE
SUBJECTS WITH VERY HIGH FALSE ALARM RATES (N=61)

	<u>CELT-S</u>	<u>CELT-L</u>	<u>CELT-L/S</u>
Test A-SC	.60	.56	.64
Test B-SC	.58	.60	.65
Test C-SC	.61	.60	.66

Test A-YN	.65	.53	.64
Test B-YN	.66	.58	.68
Test C-YN	.68	.58	.69

Table V shows that when the five subjects with very high, anomalous False Alarm Rates are culled from the group, the correlation among the three vocabulary tests scored as simple-checklist tests does not change meaningfully (rows 1, 2, 3). However, the correlation among the three vocabulary tests scored as YES/NO tests and the CELT subtests does change meaningfully (rows 4, 5, 6). The correlation becomes more consistent and stronger among the three vocabulary tests. However, does not become consistently or meaningfully

stronger than the correlation among the three vocabulary tests scored as simple-checklist tests and the CELT subtests.

Thus, these results do not support the hypothesis. Scores on the three vocabulary tests scored as a YES/NO test did not correlate more consistently and more strongly with the CELT subtests than the three vocabulary tests scored as simple-checklist tests. This was true even when the five subjects with very high anomalous False Alarm Rates were edited from the group.

Hypothesis 2

The correlation between vocabulary test scores and CELT subtest scores will be meaningfully different for subjects with 'high' False Alarm Rates (greater than 10%) versus subjects with 'low' False Alarm Rates (less than or equal to 10%) for simple-checklist scores but not for YES/NO scores.

Table VI shows the distribution of Test C False Alarm Rates for all 66 subjects.

TABLE VI
DISTRIBUTION OF TEST C FALSE ALARM RATES
FOR ALL SUBJECTS (N=66)

<u>Test C False Alarm Rate</u>	<u>Number of Subjects</u>
0 - 10%	36
11 - 20%	16
21 - 30%	4
31 - 40%	4
41 - 50%	2
51 - 60%	3
61 - 70%	0
71 - 80%	0
81 - 90%	1
91 - 100%	0

Table VI shows that the distribution of False Alarm Rates is highly skewed in the direction of low False Alarm Rates. The False Alarm rate of 10% roughly divides the entire group along a 55%-45% split. Thirty-six subjects had a Test-C False Alarm Rate less than or equal to 10%. Thirty subjects had a Test-C False Alarm Rate greater than 10%.

Table VII shows the correlation among the CELT subtest scores and vocabulary test scores for the group of subjects with 'low' False Alarm Rates--less than or equal to 10%. (The complete correlation matrix is presented as Appendix H.)

TABLE VII

CORRELATION AMONG CELT SUBTEST SCORES AND
VOCABULARY TEST SCORES FOR SUBJECTS WITH TEST C
FALSE ALARM RATES LESS THAN OR EQUAL TO 10% (N=36)

	<u>CELT-S</u>	<u>CELT-L</u>	<u>CELT-L/S</u>
Test A-SC	.67	.67	.73
Test B-SC	.65	.61	.69
Test C-SC	.67	.65	.72
Test A-YN	.68	.68	.73
Test B-YN	.67	.60	.69
Test C-YN	.68	.65	.73

TABLE VIII

CORRELATION AMONG CELT SUBTEST SCORES AND
VOCABULARY TEST SCORES FOR SUBJECTS WITH
TEST C FALSE ALARM RATES GREATER THAN 10% (N=30)

	<u>CELT-S</u>	<u>CELT-L</u>	<u>CELT-L/S</u>
Test A-SC	.56	.44	.55
Test B-SC	.64	.63	.68
Test C-SC	.63	.56	.64
Test A-YN	.24	.16	.22
Test B-YN	.72	.64	.73
Test C-YN	.57	.47	.56

The correlation shown in Table VII is very consistent among the three vocabulary tests and two scoring procedures. If there are meaningful differences, they are accounted for by the different CELT subtests. For example, the correlation among all three vocabulary tests scored both ways and CELT Structure subtests scores (column 1) range from .65 to .68. The correlation among all three vocabulary tests scored both ways and the CELT-L/S scores range from .69 to .73--somewhat higher, but still a very narrow range (column 3).

There are two somewhat anomalous relationships shown in Table VII. One is the correlation between Test B scored as a simple-checklist test and CELT Listening subtest scores (.61) (row 2, column 2). The other somewhat anomalous correlation shown in Table VII is between Test B scored as a YES/NO test and the CELT Listening subtest scores (.60) (row 5, column 2).

Table VIII shows the correlation for the remainder of the subjects, those with 'high' Test-C False Alarm Rates--greater than 10%. (The complete correlation matrix is shown as Appendix I.) For this grouping of subjects, the method used to score the vocabulary tests results in a large difference in the correlation among the vocabulary tests and CELT subtest scores. Correlation among the vocabulary tests and CELT subtests are clearly stronger and more consistent among all three tests when the vocabulary tests are scored as simple-checklist tests than as YES/NO tests.

Research Question 2

What is the optimal length for a YES/NO vocabulary test?

Hypothesis 3

Scores on Test C will correlate more consistently and more strongly with CELT subtest scores than will scores on Test A and Test B for both vocabulary test scoring procedures--the simple-checklist procedure and the YES/NO procedure.

Research Question 3

Can the level and type of L2 proficiency needed by EAP students in order to exit EAP programs be partially described in terms of necessary knowledge of the most-frequent 5,000 word families of English?

Hypothesis 4

Subjects who score above recommended EAP program exit scores on the CELT subtests (CEL_T Listening=.80; CEL_T - L/S=.73) will also score above 5,000 on the vocabulary tests.

The previous section showed that the most consistent correlation among CELT subtest scores and vocabulary test scores resulted from scoring the vocabulary tests as simple checklist tests rather than as YES/NO tests. It was also shown that the most consistent correlation was among Test C and CELT L/S scores. Thus, correlation between scores on Test C scored as a simple checklist and CELT-L/S and CELT-L scores was plotted for this analysis. Figure 1 shows the

correlation between scores on Test C scored as a simple checklist test and CELT-L/S scores for all 66 subjects.

Figure 1. Correlation between scores on Test C scored as a simple checklist test (TESTC) and CELT - L/S scores (CELTLS) for all subjects (N=66) ($r=.67$).

The correlation between Test C scores and the CELT - L/S scores was moderate ($r=.67$). The recommended cut-off points for CELT scores (.73) and also for the hypothesized 5,000 base-word threshold level for vocabulary size have been drawn. Figure 1 shows that three subjects (5% of the 66 subjects) produced a score that does not support the hypothesis. These three subjects had CELT-L/S scores above .73 but also had scores on Test C (scored as a simple checklist test) below 5,000.

Figure 2 shows the correlation between CELT Listening subtest scores and scores on Test C scored as a simple checklist for all 66 subjects.

Figure 2. Correlation between CELT Listening subtest scores (CELTL) and scores on Test C scored as a simple checklist test (TESTC) for all subjects (N=66) ($r=.61$).

Figure 3 shows that one subject (2% of the 66 subjects) produced a score that does not support the hypothesis. This subject scored above 80% on the Listening subtest but below 5,000 on the vocabulary test.

CHAPTER V

DISCUSSION

Research Question 1

Is the YES/NO test procedure a more valid measure of L2 vocabulary knowledge than the simple-checklist procedure?

Hypothesis 1

Scores on Test A, Test B, and Test C will correlate more strongly and more consistently with CELT Structure subtest scores and CELT Listening subtest scores when the vocabulary tests are scored using the YES/NO procedure than when scored as simple-checklist tests.

This hypothesis was not supported. Correlation among the vocabulary tests and CELT subtests for all subjects were more consistently stronger when the vocabulary tests were scored as simple-checklist tests than when scored as YES/NO tests. When the five subjects with very high False Alarm Rates were culled from the group, some of the correlation among the vocabulary tests and the CELT subtests became stronger when the vocabulary tests were scored as YES/NO tests than as simple-checklist tests, but not meaningfully stronger.

It seems that the YES/NO procedure is not improving the validity of the vocabulary test scores even when the subjects with very high False Alarm Rates are edited from the data (Table V). It seems that there is a relationship between subjects' False Alarm Rates and the percentage of real words that they mark but do not know. However, the adjustment formula does not seem to account for this relationship. This will be discussed further below.

Hypothesis 2

The correlation among vocabulary test scores and CELT subtest scores for subjects with 'high' False Alarm Rates (greater than 10%) versus subjects with 'low' False Alarm Rates (less than or equal to 10%) will be meaningfully different for simple checklist scores but not for YES/NO scores.

This hypothesis was not supported. The correlation was meaningfully different for both vocabulary test scoring procedures between the subgroup of subjects with low False Alarm Rates (Table VI) and the subgroup of subjects with high False Alarm Rates (Table VII). The correlation among the vocabulary tests scored both ways and the CELT subtests for subjects with 'high' False Alarm Rates was weaker and less consistent than for subjects with low False Alarm Rates. Subjects with high False Alarm Rates must be marking more real words that they do not know than subjects with low False

Alarm Rates. The problem seems to be that the adjustment formula does not adequately describe the relationship between False Alarm Rate and Miss Rate--'Miss Rate' being the percentage of real words marked but not actually known. Perhaps a different formula would work better; but on the other hand, the relationship may be highly variable from subject to subject or from test to test.

One way to explore the relationship between False Alarm Rates and Miss Rates would be to give a YES/NO test and a subsequent interview test of the real words. My feeling, however, is that this relationship is complex and variable, and that exploring it would not be the most productive type of research for the improvement of vocabulary tests. A better direction seems to be to explore how to elicit cautious performances on simple-checklist tests.

The strongest and most consistent correlation among vocabulary test scores and CELT subtest scores was found for the subgrouping of subjects with False Alarm Rates less than or equal to 10% (Table VII). For this subgrouping of subjects there was no meaningful difference in the correlation among the vocabulary tests and the CELT subtests due to the way the vocabulary tests were scored. Thus, when subjects are cautious, checklist test results become more valid. Not only was the correlation among the vocabulary tests and CELT subtests for subjects with low False Alarm Rates the strongest in this study, but it was also stronger

than that commonly found between vocabulary tests and listening and grammar tests. For the group of subjects with low False Alarm Rates, correlation of .67 was found between scores on Test C scored as a simple-checklist test and CELT Structure subtest scores. Henning and Cascallar (1992) found correlation between the 'grammar' and 'vocabulary' subcomponents of the TOEFL of .46 for a similar group of subjects. Also, for the group of subjects with low False Alarm Rates, correlation of .65 was found between scores on Test C scored as a simple-checklist test and CELT Listening subtest scores. Henning and Cascallar (1992) found correlation between the 'listening' and 'vocabulary' subcomponents of the TOEFL of .48 for a similar group of subjects. This suggests that under certain conditions checklist tests may be able to measure L2 vocabulary knowledge in a way that short multiple-choice tests cannot.

Research Question 2

What is the optimal length for a YES/NO vocabulary test?

Hypothesis 3

Scores on Test C will correlate more consistently and more strongly with CELT subtest scores than will scores on Test A and Test B for both vocabulary test scoring procedures--the simple-checklist procedure and the YES/NO procedure.

This hypothesis was supported. The correlation between vocabulary test scores and CELT subtest scores were almost always higher for Test C scores than for Test A and Test B scores. In the cases where correlation between Test A or Test B scores and CELT subtest scores was higher than the correlation between Test C scores and CELT subtest scores, the differences were not meaningful (Tables IV, V, VI, VII). Moreover, all of the somewhat anomalous correlation was between either Test A or Test B scores and the CELT subtest scores. In other words, Test C was more reliable and therefore more valid than Test A or Test B.

The important tradeoff tested in this hypothesis is that between the better confidence interval provided by a larger sample size and the potential for reduced subject performance as the test becomes longer. In other words, a longer test is better from a statistical point of view, but potential improvements to test validity in statistical terms may be canceled out by a reduction in subject performance on the test. As a subject becomes physically tired or simply bored with the task, his or her level of cautiousness in answering items may decrease substantially.

The results of this study, however, suggest that the longer 120-real-word test (Test C) is definitely a more valid test than either of the shorter 60-real-word tests (Test A and Test B). It seems reasonable then to posit that if the 60 nonsense words in Test C were replaced with real words,

the validity of the test would increase. However, since Test C takes only about ten minutes to complete, perhaps it would be interesting to evaluate even longer tests.

Research Question 3

Can the level of L2 proficiency needed by EAP students in order to exit ESL programs be partially described in terms of necessary knowledge of the most-frequent 5,000 word families of English?

Hypothesis 4

Subjects who score above recommended cut-off scores on the CELT subtests (CELT Listening=.80; CELT Listening/Structure avg.=.73) will also score above 5,000 on the vocabulary test.

This hypothesis was supported. Only 5% of the subjects had CELT-L/S scores above .73 and Test C scores below 5,000 (Figure 1). Only 2% of the subjects scored above 80 on the CELT Listening subtest but below 5,000 on the vocabulary test (Figure 2). While there is some evidence provided here that knowledge of the most-frequent 5,000 word families of English is necessary to obtain exit-level scores on the CELT subtests, there is no evidence nor claim made that such knowledge is sufficient. Fourteen percent of the subjects scored above 5,000 on Test C but had a CELT-L/S score below the recommended cut-off point. Eighteen percent of the

subjects scored above 5,000 on Test C but had a CELT Listening subtest score below the recommended cut-off point.

Several questions are raised by these results. Are the students who score above 5,000 on the vocabulary test (Test C scored as a simple-checklist test) but below recommended EAP program exit scores on the CELT subtests studying too much vocabulary? Should they turn their attention to studying grammar or to listening practice? Are they good readers but poor at listening? These questions cannot be answered from the data provided by this study. However, it is suggested by the results of this study that some of the subjects that scored above 5,000 on the vocabulary test (Test C) but below the recommended cut-off point on the CELT Listening subtest are better at grammar than at listening. When these subjects' CELT Structure subtest scores are averaged with their not-passing CELT Listening subtest scores, 25% of these subjects obtain a CELT-L/S score that is above the recommended cut-off point. Moreover, CELT-L/S scores are merely an approximation of what a subject would score if that subject had taken the complete CELT test.

The third and final section, which unfortunately was not able to be given, is a 75-item vocabulary test. If the subjects with vocabulary test scores (Test C) above 5,000 but CELT-L/S scores below the cut-off point had taken the 'vocabulary' subtest of the CELT, it seems likely that their performances on this subtest would have pushed their actual

total CELT scores above the recommended cut-off point. In future studies it would be interesting to look at the relationship between scores on a checklist vocabulary test and a more extensive battery of L2 proficiency tests.

IMPLICATIONS AND SUGGESTIONS FOR FUTURE RESEARCH

The main implications of this study are for vocabulary testing. By comparing the vocabulary tests, scored as both a simple-checklist tests and as YES/NO tests, it was concluded that the YES/NO scores were not more valid than simple-checklist scores. This conclusion is at odds with the conclusions of all of the previous YES/NO-test validity studies reviewed earlier. However, the results of this study are not inconsistent with the results of the YES/NO-test validity studies reviewed earlier. All of the studies reviewed concluded that the YES/NO method is a good, valid method. However, all of the YES/NO-test validity studies only compared the strength of correlation among YES/NO vocabulary tests, multiple-choice vocabulary tests, oral-interview vocabulary tests, and grammar and listening comprehension tests. It was found that YES/NO vocabulary tests seemed better than multiple-choice vocabulary tests at predicting performance on reading and listening comprehension tests and interview vocabulary tests. However, none of the YES/NO-test validity studies compared YES/NO test results with simple-checklist test results. The results of this

study suggest that the YES/NO tests (which proved to be more valid measures of vocabulary size than multiple-choice tests) would have been shown more valid than the multiple-choice tests even if they had been scored as simple-checklist tests.

The most promising direction then for the improvement of vocabulary tests intended for some types of SLA research is not the improvement of multiple-choice tests or the improvement of the YES/NO test adjustment formula. Rather, the results of this study suggest that the most promising direction for the improvement of vocabulary test procedures to be used in studying the second language acquisition of adult EAP students lies in the direction of improving simple-checklist tests.

One way to do this might be to carefully design the test instructions and to carefully control how the test is given. One possible method is to give a simple-checklist vocabulary test as a self-assessment. The test instructions could explain how the test was constructed, what it measures, why this information is important, and so on. In other words, the test instructions could be written with the goal of persuading the subjects to want to know the information provided by the test. Hopefully, test instructions of this type would influence students to complete the test carefully and accurately.

The instructions for the test developed in this thesis simply asked students to "circle the words that you know."

However, word knowledge exists on a continuum. Some recent researchers using vocabulary tests have tried to address this issue with the use of 'vocabulary knowledge scales.' The results of these studies show promise for this technique. A vocabulary knowledge scale is a Likert scale with the points along the scale defined in terms of how well a word is known.

The idea that vocabulary acquisition and vocabulary knowledge exist on a continuum that can be pragmatically reduced to 'stages' is not new. In 1965, Dale proposed the following four stages of word knowledge:

Stage 1: "I never saw it before."

Stage 2: "I've heard of it, but I don't know what it means."

Stage 3: "I recognize it in context--it has something to do with..."

Stage 4: "I know it".

Several recently developed vocabulary knowledge scales appear to have been developed independently and also without influence from Dale (1965) since no references are made between most of the studies or to Dale's work. Therefore, it is interesting that the scales are quite parallel.

D'Anna et al. (1991), for instance, constructed the following scale:

- 1 = "I have never experienced the word before."
- 2 = "I have seen or heard the word before, but I do not know its meaning."
- 3 = "I have seen or heard the word before and have a vague idea of its meaning."
- 4 = "I would be able to recognize the meaning of the word if given a multiple-choice test which included the correct meaning and several incorrect meanings."
- 5 = "I know the meaning of this word well enough to give its definition."

Stages 1, 2, and 3 of D'Anna et al.'s and Dale's scales are exactly parallel. The only difference between the two scales is that the D'Anna scale has two stages for certain knowledge while Dale's scale has only one--Stage 4. D'Anna's Stage 4 corresponds generally to the concept of 'receptive' knowledge while Stage 5 corresponds generally to the concept of 'productive' knowledge.

Paribakht and Wesche (1993) constructed and used the following scale:

- I. "I have never seen this word."
- II. "I have seen this word before, but I don't know what it means."
- III. "I have seen this word before, and I think it means _____. (synonym or translation)"
- IV. "I know this word. It means _____. (synonym or translation)"
- V. "I can use this word in a sentence: _____."

Paribakht and Wesche's scale is very similar to D'Anna's and Dale's scales. It differs mainly in that it asks subjects not to only choose a stage but also to demonstrate the knowledge which defines that stage. This technique worked for Paribakht and Wesche since they used the scale to test only a small number of words. D'Anna, on the other hand, tested a large number of words, making the inclusion of translations or synonyms infeasible. Goulden et al. also tested a large number of words and were influenced by the work of D'Anna et al.

Goulden et al. (1990) developed and used the following scale:

- Stage 1 = "I don't know this word."
- Stage 2 = "I think I know this word, but I'm not sure."
- Stage 3 = "I know this word."

This seems like a streamlined version of the previous scales. Stages 1 and 2 of the previous scales are collapsed into one stage in the Goulden scale. Stage 2 in the Goulden scale corresponds almost perfectly to Stage 3 of the other scales. Stage 3 in the Goulden scale corresponds exactly to Dale's Stage 4 and is a collapsed version of Stages 4 and 5 in the other two previous scales. The following scale, which is very much parallel to the Goulden scale, was used with high school students by Dole et al. (1995):

Stage 1	=	"unknown"
Stage 2	=	"acquainted"
Stage 3	=	"established"

McKenzie (1990) did not use the following scale in a test but seems to have created it without influence from any of the scale developers mentioned above:

Stage 1	= "ZONE OF TOTALLY UNKNOWN WORDS Words that you have never heard or seen before."
Stage 2	= "INTERMEDIATE ZONE Words that you partially understand."
Stage 3	= "ZONE OF KNOWN WORDS Words that you know well."

Three of the recent studies that used vocabulary knowledge scales used them in an identical manner. Subjects were asked to indicate their knowledge of specific words according to the scale, but when the researchers analyzed the data, words were classified as either 'known' or 'unknown' using some cut-off point on the scale. For D'Anna et al. and Paribakht and Wesche, the cut-off point was located between Stages 3 and 4. Thus, if a subject marked a word '1,' '2,' or '3,' the word was classified as 'unknown.' If a subject marked a word '4' or '5,' the word was classified as 'known.' For Goulden et al., the cut-off point was between Stage 2 and Stage 3. Therefore, there is some agreement that even if a subject has seen and has a vague idea of a word's meaning,

that word should still be classified as 'unknown' for practical purposes.

If subjects are asked to mark each word in a test with a number from a vocabulary knowledge scale, it will take much more time to complete the test than if the simple-checklist procedure is used. Perhaps the best way to use such scales is to use them as directions for when to 'check' or 'not check' a word. For example, checklist test directions could say something to the effect that "if your knowledge of a specific word is at Stage 1, 2, or 3, do not check the word; if your knowledge of the word is at stages 3 or 4, check the word."

A second way to improve simple-checklist tests might be to include some nonsense words in the test even though the YES/NO scoring procedure is not used. This could give researchers confidence in the validity of their results, and subjects with high False Alarm Rates might be edited from group analyses.

A third way that checklist vocabulary tests intended for adult EAP students could be improved is by continuing to refine samples of test words. The results of this study indicate that a test longer than 120 real words might be more useful. Perhaps a 250- or 300-word test would be possible. A simple checklist test containing 250-300 words would take most adult EAP students only about twenty minutes to complete. Also, the most-frequent 1,000 word families could

be eliminated from the source list. Results of this study indicate that knowledge of these words can be assumed for nearly all EAP students. This would allow the remaining portion of the source list to be sampled at a higher rate using fewer test words. If Francis and Kucera's (1982) adjusted-rank lemma frequency list were used, excluding the first 1,000 words would leave a 5,000-word list. A 250-word sample of this 5,000-word list would produce a test with a minimum 95% confidence interval of plus or minus 300 source-list words. This means that, whatever the subject's score on the test, it could be said with 95% confidence that the subject's true score is at most within the range of plus or minus 300 source-list words from that subject's observed score. In statistical terms, this would be a big improvement over Test C, which has a minimum 95% confidence interval of plus or minus 540 source-list words. (Confidence intervals for various size tests sampled from a 5,000-word source list are shown in Table II.)

Such a checklist test has been created and appears as Appendix J. It was created by randomly sampling one out of every twenty words in Francis and Kucera's (1982) adjusted-rank lemma-frequency list. The first 1,000 words were excluded. The test words are randomly listed in frequency bands such that the first band lists words from the second 1,000 most-frequent lemmas, the second band list words from the third 1,000 most-frequent lemmas, and so on. Ten

nonsense words are randomly listed with each frequency band, not to be used in YES/NO scoring but just to give researchers an idea of the level of cautiousness used by their subjects. This test has been pilot-tested with intermediate-level, adult EAP students and takes about twenty minutes to complete.

The test is included here for two reasons. First, it is an example of one way that future checklist tests can be constructed. Such parallel tests could be useful measures of vocabulary growth in long-term method-comparison studies, for example. Second, the test is believed to be a practical diagnostic test.

This test can be scored in two ways. One way is just to calculate the percentage of real words checked. If a score that reflects number of word families known is desired, the percentage of real words checked can be multiplied by 5,000. The resulting figure can be added to 1,000 to give an estimate of the number of the most-frequent 6,000 lemmas in written English known.

It would be interesting to see how this test compares with other diagnostic tests such as Nation's (1990) 'vocabulary levels test' and how scores on this test correlate with various tests of L2 proficiency using various types of test instructions.

CONCLUSIONS

The results of this study suggest that the simple-checklist procedure, when combined with the procedure of sampling test words from a word-family frequency list, can be a useful way to test the lexical knowledge of adult EAP students. The construction and use of such tests may be a way to diagnose the needs of specific students and may be useful in several kinds of studies.

Simple-checklist vocabulary tests show promise as a way to continue to explore the relationship between vocabulary knowledge and various aspects of L2 ability. This study, for example, has provided support for the hypothesis that knowledge of the 5,000 most-frequent word families of English is necessary for the acquisition of the level and type of L2 ability required to exit EAP programs. Simple-checklist tests might also be useful in long-term methods-comparison studies.

Simple-checklist vocabulary tests might also be useful to EAP students in direct ways. Most students are aware of their need for more vocabulary knowledge, but few are aware of how many words they need to know or do know. Also, if the test directions use a vocabulary knowledge scale, the test may help students learn to evaluate the depth as well as the breadth of their vocabulary knowledge.

REFERENCES

- American Association of College Registrars and Admissions Officers. (1994). The Guide: A resource for international admission professionals Washington, DC: Author.
- Anderson, R. C., & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. In B. A. Hutson (Ed.), Advances in reading/language research 2 (pp. 231-256). Greenwich, CT: JAI Press.
- Bright, J.A., & McGregor, G.P. (1970). Teaching English as a second language. London: Longman.
- Carver, R.P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: implications for instruction. Journal of Reading Behavior, 26(4), 413-437.
- Cooksey, R. W., & Freebody, P. (1987). Aspects of a computer-managed test of children's reading vocabulary: reliability, validity, and characterization of knowledge. Reading Psychology, 8, 103-118.
- Dale, E. (1965). Vocabulary measurement: Techniques and major findings. Elementary English, 42, 895-901, 948.
- D'Anna, C.A., Zechmeister, E.B., & Hall, J.W. (1991). Toward a meaningful definition of vocabulary size. Journal of Reading Behavior, 23(1), 109-122.
- Dole, J.A., Sloan, C., & Trathen, W. (1995). Teaching vocabulary within the context of literature. Journal of Reading, 38(6), 452-460.
- Ellegard, A. (1960). Estimating vocabulary size. Word, 16(2), 219-244.
- Ferris, D. R. (1991). Syntactic and lexical characteristics of ESL student writing: A multidimensional study. Unpublished doctoral dissertation, University of Southern California.

- Ferris, D.R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. TESOL Quarterly, 28(2), 414-420.
- Fox, L. (1979). On acquiring an adequate second language vocabulary. Journal of Basic Writing, 2(3), 68-75.
- Francis, W. N., & Kucera, H. (1982). Frequency analysis of English usage: Lexicon and grammar. Boston: Houghton Mifflin.
- Gass, S.M., & Selinker, L. (1994). Second language acquisition: An introductory course. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gorman, T.P. (1979). Teaching reading at the advanced level. In M. Celce-Murcia & L. McIntosh (Eds.) Teaching English as a second or foreign language. pp. 154-161. Rowley, MA: Newbury House.
- Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? Applied Linguistics, 11(4), 341-363.
- Grabe, W. (1991). Current developments in second language reading research. TESOL Quarterly, 25(3), 375-406.
- Graves, M. F., Ryder, R. J., Slater, W. H., & Calfee, R. C. (1987). The relationship between word frequency and reading vocabulary using six metrics of frequency. Journal of Educational Research, 81(2), 81-90.
- Harris, D. P., & Palmer, L. A. (1986). A comprehensive English language test for learners of English: Examiner's instructions and technical manual. New York: McGraw-Hill.
- Hayes, D.P., & Ahrens, M.G. (1988). Vocabulary simplification for children: A special case of 'motherese'? Journal of Child Language, 15, 395-410.
- Henning, G., & Cascallar, E. (1992). A preliminary study of the nature of communicative competence. TOEFL Research Report #36. Princeton, NJ: Educational Testing Service.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? Reading in a Foreign Language, 8(2), 689-696.

- Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. IRAL, 29(2), 135-149.
- Kucera, H., & Francis, W. N. (1967). A computational analysis of present-day American English. Providence, RI: Brown University Press.
- Kyongho, H., & Nation, P. (1989). Reducing the vocabulary load and encouraging vocabulary learning through reading newspapers. Reading in a Foreign Language, 6(1), 323-335.
- Laufer, B. (1989). What percentage of text-lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), Special language: From humans thinking to thinking machines (pp. 316 - 323). Clevedon: Multilingual Matters.
- Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Bejoint (Eds.), Vocabulary and applied linguistics (pp. 126-132). London: Macmillan.
- Leki, I., & Carson, J. G. (1994). Students' perceptions of EAP writing instruction and writing needs across the disciplines. TESOL Quarterly, 28(1), 81-101.
- Liu, N., & Nation, P. (1985). Factors affecting guessing vocabulary in context. RELJ Journal, 16(1), 33-42.
- McKenzie, M. (1990). Letting lexis come from the learner: A word in the hand is worth two in the bush. English Teaching Forum, 21(1), 13-16.
- Meara, P. (1992). EFL vocabulary tests. (ERIC Document Reproduction Service No. ED 362 046).
- Meara, P., & Buxton, B. (1987). An alternative to multiple choice tests. Language Testing, 4(2), 142-151.
- Meara, P., & Jones, G. (1988). Vocabulary size as a placement indicator. (ERIC Document Reproduction Service No. ED 350 829).
- Meara, P., Lightbown, P. M., & Halter, R. H. (1994). The effect of cognates on the applicability of YES/NO vocabulary tests. Canadian Modern Language Review, 50(2), 296-311.
- Nation, I. S. P. (1990). Teaching and learning vocabulary. New York: Newbury House.

- Nation, P. (1993). Using dictionaries to estimate vocabulary size: essential but rarely followed procedures. Language Testing, 10(1), 27-40.
- Nation, P. (Ed.) (1994). New ways in teaching vocabulary. Alexandria, VA: Teachers of English to Speakers of Other Languages.
- Ostyn, P., & Godin, P. (1985). RALEX: An alternative approach to language teaching. Modern Language Journal, 69(4), 346-355.
- Paribakht, T.S., & Wesche, M.B. (1993). Reading comprehension and second language development in a comprehension-based ESL program. TESL Canada Journal, 11 (1), 9-29.
- Read, J. (1988). Measuring the vocabulary knowledge of second language learners. RELC Journal, 19(2), 12-25.
- Ryder, R.J., & Slater, W.H. (1988). The relationship between word frequency and word knowledge. Journal of Educational Research, 81(5), 312-317.
- Saville-Toikey, M. (1984). What really matters in second language learning for academic achievement? TESOL Quarterly, 18(2), 199-219.
- Sloan, C. B. (1994). Planning for academic success: Survey of university professors' assessments of non-native students' language skill needs. Unpublished master's thesis, Portland State University, Portland, OR.
- Strother, J. B., & Ulijn, J. M. (1987). Does syntactic rewriting affect English for science and technology (EST) text comprehension. In J. Devine, P. L. Carrell, & D. E. Eskey (Eds.), Research in reading in English as a second language (pp. 91-101). Washington, D.C.: Teachers of English to Speaker of Other Languages.
- Thorndike, E. L., & Lorge, I. (1944). The teacher's word book of 30,000 words. New York: Teachers College Columbia University.
- Ulijn, J. M. (1984). Reading for professional purposes: Psycholinguistic evidence in a cross-linguistic perspective. In A. K. Pugh & J. M. Ulijn (Eds.), Reading for professional purposes: Studies and practices in native and foreign languages (pp. 66-81). London: Heinemann.

White, T. G., Slater, W. H., & Graves, M. F. (1989). YES/NO method of vocabulary assessment: Valid for whom and useful for what? Cognitive and social perspectives for literacy research and instruction: Thirty-eighth yearbook of the national reading conference (pp. 391-397). Chicago, IL: National Reading Conference.

Zimmerman, J., Broder, P. K., Shaughnessy, J. J., & Underwood, B. J. (1977). A recognition test of vocabulary using signal-detection measures, and some correlates of word and nonword recognition. Intelligence, 1, 5-31.

APPENDIX A

TEST A

1. half	24. battery	47. compare	70. accordance
2. feld	25. convenient	48. oppensive	71. perhaps
3. vigorously	26. hap	49. delectively	72. effective
4. theological	27. tremble	50. own	73. consequence
5. chip	28. ancestor	51. lurge	74. coltivate
6. company	29. deplete	52. keln	75. wift
7. either	30. plead	53. ardacy	76. broaden
8. respectively	31. queen	54. trectate	77. wool
9. department	32. formerly	55. hance	78. recorisation
10. aide	33. correlation	56. bespect	79. calculation
11. brice	34. solution	57. fintion	80. remonistrate
12. plate	35. fourth	58. beneath	81. stimulate
13. incution	6. acquisition	59. eighteen	82. prostrate
14. bond	37. unchanged	60. inquire	83. lecture
15. plense	38. lay	61. sustive	84. evil
16. danger	39. incentive	62. prodacit	85. conservation
17. tin	40. far	63. wisdom	86. physics
18. basis	41. damp	64. dislution	87. fisherman
19. snap	42. loose	65. prodiment	88. solcate
20. recontative	43. tissue	66. deepest	89. opening
21. forpel	44. nose	67. reventive	90. orthodox
22. compel	45. lomulate	68. park	
23. folk	46. colleague	69. scand	

APPENDIX B

TEST B

91. stipel	114. educate	137. woman	160. strongest
92. preventive	115. privacy	138. disorder	161. fordally
93. prosectify	116. comparative	139. pistently	162. arstance
94. prove	117. listure	140. reassure	163. recentive
95. candigent	118. vasculogy	141. physiture	164. declaration
96. mirror	119. messenger	142. scare	165. profigate
97. acknowledge	120. consile	143. ago	166. drast
98. outside	121. execution	144. locate	167. thrute
99. vast	122. thrust	145. brick	168. warrant
100. excind	123. useless	146. price	169. keen
101. convict	124. suspect	147. schedule	170. production
102. bee	125. candidate	148. ordantly	171. quickly
103. sustend	126. render	149. displacement	172. hay
104. resistance	127. singular	150. basend	173. lesson
105. socation	128. revict	151. arrome	174. conscience
106. finally	129. little	152. ainge	175. sir
107. arbitrary	130. unlike	153. gifted	176. tald
108. soduction	131. amazement	154. task	177. vastulent
109. prode	132. example	155. steen	178. necessity
110. accentary	133. newspaper	156. expedition	179. professor
111. precessity	134. onset	157. negative	180. become
112. commodity	135. expectation	158. arouse	
113. depending	136. depellicant	159. classic	

APPENDIX C

RANK OF TEST WORDS IN FRANCIS AND KUCERA'S (1982)

ADJUSTED-RANK LEMMA FREQUENCY LIST

Test A			Test B		
#	Real Words	Rank	#	Real Words	Rank
1.	own	100	91.	become	99
2.	company	200	92.	woman	199
3.	perhaps	300	93.	example	299
4.	half	400	94.	ago	399
5.	department	500	95.	finally	499
6.	basis	600	96.	prove	599
7.	lay	700	97.	price	699
8.	far	800	98.	production	799
9.	effective	900	99.	little	899
10.	compare	998	100.	newspaper	999
<hr/>					
11.	park	1100	101.	quickly	1099
12.	danger	1200	102.	task	1199
13.	solution	1300	103.	locate	1299
14.	fourth	1398	104.	professor	1399
15.	either	1500	105.	vast	1499
16.	consequence	1600	106.	outside	1599
17.	bond	1700	107.	candidate	1699
18.	folk	1800	108.	necessity	1799
19.	beneath	1900	109.	suspect	1899
20.	nose	2000	110.	lesson	1999
<hr/>					
21.	queen	2100	111.	unlike	2098
22.	loose	2200	112.	resistance	2199
23.	plate	2300	113.	render	2299
24.	wisdom	2400	114.	schedule	2399
25.	opening	2500	115.	negative	2498
26.	lecture	2598	116.	conscience	2597
27.	tissue	2700	117.	educate	2699
28.	evil	2800	118.	arouse	2799
29.	snap	2898	119.	expectation	2899
30.	tremble	3000	120.	acknowledge	2999
<hr/>					
31.	colleague	3100	121.	sir	3099
32.	formerly	3300	122.	mirror	3199
33.	inquire	3300	123.	brick	3299
34.	compel	3400	124.	classic	3399
35.	physics	3500	125.	declaration	3499
36.	plead	3600	126.	depending	3599
37.	battery	3698	127.	execution	3699
38.	broaden	3800	128.	thrust	3799
39.	respectively	3900	129.	strongest	3899
40.	stimulate	4000	130.	bee	3999

41.	convenient	4098	131.	arbitrary	22	4098
42.	calculation	4200	132.	useless	17	4199
43.	accordance	4300	133.	scare	6	4298
44.	eighteen	4400	134.	commodity	28	4398
45.	theological	4500	135.	hay	19	4499
46.	damp	4600	136.	expedition	21	4599
47.	acquisition	4700	137.	displacement	23	4699
48.	orthodox	4800	138.	onset	43	4798
49.	vigorously	4899	139.	gifted	13	4899
50.	ancestor	5000	140.	singular	14	4999

51.	incentive	5099	141.	convict	16	5098
52.	aide	5199	142.	warrant	14	5198
53.	correlation	5300	143.	comparative	17	5297
54.	tin	5399	144.	preventive	15	5393
55.	conservation	5500	145.	disorder	14	5498
56.	fisherman	5600	146.	privacy	12	5595
57.	unchanged	5699	147.	messenger	12	5697
58.	wool	5799	148.	amazement	10	5799
59.	deepest	5988	149.	reassure	12	5899
60.	chip	5996	150.	keen	11	5993

Test A - Nonsense Words

61. plense
62. wift
63. prost
64. brice
65. hap
66. marl
67. fane
68. scand
69. feld
70. prostrate
71. surrent
72. lithic
73. displore
74. fintion
75. solcate
76. deple
77. mascent
78. forpel
79. lomulate
80. prodiment
81. prodacit
82. reventive
83. incution
84. oppensive
85. ardacy
86. dislution
87. rementative
88. remonistrate
89. delectively
90. recorisation

Test B - Nonsense Words

151. mant
152. swant
153. jure
154. drast
155. crask
156. prode
157. stipel
158. excind
159. sustend
160. consile
161. arrome
162. intone
163. rilent
164. listure
165. lective
166. accentary
167. symetic
168. candigent
169. soduction
170. absestate
171. redition
172. fordally
173. ordantly
174. profigate
175. recentive
176. elecism
177. precessity
178. dismonital
179. releity
180. prosectify

Note:

- (1) 'Rank' = ranking of the word in the adjusted-rank lemma frequency list from Francis and Kucera (1982)
- (2) '#' does not reflect the order of the words in the words in the vocabulary tests

APPENDIX D

VOCABULARY TEST INSTRUCTIONS

Vocabulary Size Test

Name: _____

First language: _____

=====

This test measures how many words that you know. It takes about 7 minutes. The test is a list of words. Some of the words are easy English words. Some of the words are difficult English words. Some of the words are not English words.

=====

DIRECTIONS:

Look at the list of words on page 2 and page 3. Circle the words that you know. Remember, some of the words are not real English words.

=====

EXAMPLE:

1.	<u>book</u>	(I know this word.)
2.	<u>house</u>	(I know this word.)
3.	burdle	(I <u>don't know</u> this word.)
4.	relve	(I <u>don't know</u> this word.)
5.	<u>dazzle</u>	(I know this word.)
6.	conversal	(I <u>don't know</u> this word.)
- =====

APPENDIX E

CORRELATION AMONG ALL THREE CELT SUBTEST SCORES AND
ALL THREE VOCABULARY TEST SCORES FOR ALL SUBJECTS (N=66)

**CORRELATION AMONG ALL THREE CELT SUBTESTS AND
ALL THREE VOCABULARY TESTS FOR ALL SUBJECTS (N=66)**

	CELT-S	CELT-L	CELTLS	A-SC	B-SC	C-SC	A-YN	B-YN	C-YN
CELT-S	1.00								
CELT-L	.69	1.00							
CELTLS	.92	.92	1.00						
A-SC	.60	.57	.64	1.00					
B-SC	.60	.61	.66	.90	1.00				
C-SC	.62	.61	.67	.97	.98	1.00			
A-YN	.44	.39	.46	.74	.58	.67	1.00		
B-YN	.68	.62	.71	.89	.95	.95	.67	1.00	
C-YN	.63	.56	.65	.91	.84	.89	.91	.92	1.00

APPENDIX F

DATA FOR ALL SUBJECTS (N=66)

Subject	CELT-S	CELT-L	CELTIs	YNAPk	YNBpk	YNCpk	YNAhr	YNBhr	YNChr
1	37	56	47	57	56	56	.600	.600	.600
2	51	60	56	74	74	74	.783	.783	.783
3	31	42	37	55	57	56	.567	.567	.567
4	31	30	31	34	30	32	.383	.367	.375
5	48	48	48	75	57	66	.750	.600	.675
6	71	92	82	88	92	90	.883	.933	.908
7	33	42	38	41	44	43	.450	.517	.483
8	60	66	63	65	73	69	.700	.783	.742
9	47	48	48	62	54	58	.667	.617	.642
10	53	52	53	62	57	59	.617	.567	.592
11	39	58	49	44	46	45	.517	.550	.533
12	52	66	59	73	73	73	.733	.750	.742
13	61	82	72	93	86	90	.933	.867	.900
14	53	62	58	72	65	69	.733	.650	.692
15	40	30	35	54	55	54	.567	.567	.567
16	61	46	54	71	94	82	.767	.950	.858
17	51	42	47	72	68	70	.750	.700	.725
18	45	40	43	56	57	56	.600	.583	.592
19	68	72	70	87	83	85	.867	.833	.850
20	67	62	65	97	96	96	.967	.967	.967
21	77	56	67	79	80	80	.800	.800	.800
22	73	78	76	78	89	83	.800	.917	.858
23	37	44	41	85	91	88	.850	.917	.883
24	69	60	65	78	82	80	.783	.817	.800
25	72	50	61	97	97	97	.967	.967	.967
26	79	80	80	67	69	68	.667	.700	.683
27	45	36	41	59	52	55	.600	.517	.558
28	20	32	26	50	50	50	.667	.667	.667
29	59	52	56	76	78	77	.783	.783	.783
30	44	30	37	48	47	47	.500	.483	.492
31	45	36	41	77	65	71	.800	.683	.742
32	60	58	59	70	61	65	.700	.633	.667
33	71	94	83	96	98	97	.967	.983	.975
34	68	64	66	83	85	84	.850	.867	.858
35	83	56	70	88	92	90	.883	.917	.900
36	51	72	62	77	67	72	.767	.700	.733
37	44	54	49	67	58	63	.667	.583	.625
38	29	18	24	40	40	40	.400	.400	.400
39	39	36	38	63	54	58	.800	.800	.800
40	51	52	52	73	69	71	.767	.700	.733
41	84	78	81	0	90	50	.883	.983	.933
42	25	48	37	71	71	71	.733	.767	.750
43	21	42	32	74	59	66	.767	.683	.725
44	52	74	63	78	70	74	.783	.717	.750
45	41	68	55	44	71	58	.667	.767	.717
46	39	52	46	78	82	80	.867	.900	.883
47	43	50	47	41	47	44	.467	.483	.475
48	52	76	64	56	59	58	.650	.767	.708
49	45	36	41	68	70	69	.733	.717	.725
50	65	74	70	67	68	67	.733	.767	.750
51	85	78	82	81	73	77	.817	.733	.775
52	61	46	54	60	55	58	.600	.550	.575
53	56	40	48	62	70	66	.633	.700	.667
54	37	56	47	40	58	49	.600	.750	.675
55	64	58	61	70	74	72	.767	.833	.800
56	75	86	81	76	80	78	.783	.850	.817
57	75	72	74	87	100	92	.933	1.00	.967
58	81	62	72	83	69	76	.850	.733	.792
59	80	80	80	92	100	97	.967	1.00	.983
60	57	48	53	72	66	69	.733	.750	.742
61	35	52	44	65	67	66	.650	.667	.658
62	61	56	59	71	67	69	.717	.683	.700
63	39	40	40	58	57	57	.633	.667	.650
64	37	42	40	58	59	58	.717	.633	.675
65	61	60	61	60	59	60	.617	.617	.617
66	75	68	72	70	72	71	.700	.717	.708

<u>Subject</u>	<u>YNAfar</u>	<u>YNBfar</u>	<u>YNCfar</u>
1	.067	.100	.083
2	.167	.167	.167
3	.033	.000	.017
4	.067	.100	.083
5	.000	.067	.033
6	.067	.200	.133
7	.067	.133	.100
8	.133	.200	.167
9	.133	.167	.150
10	.000	.000	.000
11	.133	.167	.150
12	.000	.067	.033
13	.000	.067	.033
14	.033	.000	.017
15	.067	.033	.050
16	.200	.200	.200
17	.100	.067	.083
18	.100	.033	.067
19	.000	.000	.000
20	.000	.133	.067
21	.033	.000	.017
22	.100	.267	.183
23	.000	.033	.017
24	.000	.000	.000
25	.033	.000	.017
26	.000	.033	.017
27	.033	.000	.017
28	.333	.333	.333
29	.100	.000	.050
30	.033	.033	.333
31	.133	.100	.117
32	.000	.067	.033
33	.067	.067	.067
34	.133	.133	.133
35	.000	.000	.000
36	.000	.100	.050
37	.000	.000	.000
38	.000	.000	.000
39	.467	.567	.517
40	.133	.033	.083
41	.900	.833	.867
42	.067	.200	.133
43	.100	.233	.167
44	.000	.067	.033
45	.400	.200	.300
46	.400	.433	.417
47	.100	.033	.133
48	.200	.433	.317
49	.167	.067	.117
50	.200	.267	.233
51	.033	.000	.017
52	.000	.000	.000
53	.033	.000	.017
54	.333	.400	.367
55	.233	.367	.600
56	.100	.267	.183
57	.500	.667	.583
58	.133	.133	.133
59	.600	.333	.467
60	.033	.267	.150
61	.000	.000	.000
62	.033	.033	.033
63	.133	.233	.183
64	.333	.100	.217
65	.033	.067	.050
66	.000	.000	.000

Key:

CELT-S =	scores on the Structure subsection of the CELT
CELT-L =	scores on the Listening subsection of the CELT
CELTls =	scores created by averaging a student's CELT-S and CELT-L scores
TSTApk =	scores on vocabulary Test A scored using the YES/NO procedure
TSTBpk =	scores on vocabulary Test B scored using the YES/NO procedure
TSTCpk =	scores on vocabulary Test C scored using the YES/NO procedure
TSTAhr =	scores on vocabulary Test A scored as a simple-checklist test
TSTBhr =	scores on vocabulary Test B scored as a simple-checklist test
TSTChr =	scores on vocabulary Test C scored as a simple-checklist test
TSTAfar =	False Alarm Rates on vocabulary Test A
TSTBfar =	False Alarm Rates on vocabulary Test B
TSTCfar =	False Alarm Rates on vocabulary Test C

APPENDIX G

CORRELATION AMONG ALL THREE CELT SUBTEST SCORES AND
ALL THREE VOCABULARY TEST SCORES FOR ALL SUBJECTS EXCEPT
FIVE SUBJECTS WITH VERY HIGH FALSE ALARM RATES (N=61)

**CORRELATION AMONG CELT SUBTESTS AND
VOCABULARY TESTS FOR ALL SUBJECTS EXCEPT FIVE
SUBJECTS WITH EXTREMELY HIGH FALSE ALARM RATES (N=61)**

	CELT-S	CELT-L	CELTLS	A-SC	B-SC	C-SC	A-YN	B-YN	C-YN
CELT-S	1.00								
CELT-L	.66	1.00							
CELTLS	.91	.91	1.00						
A-SC	.60	.56	.64	1.00					
B-SC	.58	.60	.65	.89	1.00				
C-SC	.61	.60	.66	.97	.97	1.00			
A-YN	.65	.53	.64	.95	.79	.90	1.00		
B-YN	.66	.58	.68	.90	.96	.96	.86	1.00	
C-YN	.68	.58	.69	.96	.91	.96	.97	.97	1.00

APPENDIX H

CORRELATION AMONG ALL THREE CELT SUBTEST SCORES AND
ALL THREE VOCABULARY TEST SCORES FOR SUBJECTS WITH TEST C
FALSE ALARM RATES LESS THAN OR EQUAL TO 10% (N=36)

**CORRELATION AMONG CELT SUBTEST SCORES AND
VOCABULARY TEST SCORES FOR SUBJECTS WITH TEST C
FALSE ALARM RATES LESS THAN OR EQUAL TO 10% (N=36)**

	CELT-S	CELT-L	CELTLS	A-SC	B-SC	C-SC	A-YN	B-YN	C-YN
CELT-S	1.00								
CELT-L	.66	1.00							
CELTLS	.91	.91	1.00						
A-SC	.67	.67	.73	1.00					
B-SC	.65	.61	.69	.95	1.00				
C-SC	.67	.65	.72	.99	.99	1.00			
A-YN	.68	.68	.73	.99	.94	.98	1.00		
B-YN	.67	.60	.69	.94	.99	.98	.94	1.00	
C-YN	.68	.65	.73	.98	.98	.99	.99	.99	1.00

APPENDIX I

CORRELATION AMONG ALL THREE CELT SUBTEST SCORES AND
ALL THREE VOCABULARY TEST SCORES FOR SUBJECTS WITH TEST C
FALSE ALARM RATES GREATER THAN 10% (N=30)

**CORRELATION AMONG CELT SUBTEST SCORES AND
VOCABULARY TEST SCORES FOR SUBJECTS WITH
TEST C FALSE ALARM RATES GREATER THAN 10% (N=30)**

	CELT-S	CELT-L	CELTLS	A-SC	B-SC	C-SC	A-YN	B-YN	C-YN
CELT-S	1.00								
CELT-L	.74	1.00							
CELTLS	.94	.93	1.00						
A-SC	.56	.44	.55	1.00					
B-SC	.64	.63	.68	.85	1.00				
C-SC	.63	.56	.64	.96	.97	1.00			
A-YN	.24	.16	.22	.54	.35	.46	1.00		
B-YN	.72	.64	.73	.83	.93	.92	.44	1.00	
C-YN	.57	.47	.56	.82	.75	.82	.85	.85	1.00

APPENDIX J

NEW CHECKLIST TEST

panel	capacity	synetic	distinction	emerge	entitle
capable	impression	consist	derive	tension	internal
consile	retain	provision	eliminate	aware	tend
region	abandon	extent	contribution	facility	marl
vast	acquire	vital	saint	proceed	resolution
arome	ancient	pace	expression	fordally	eventually
crack	urge	emphasis	intensity	disple	prospect
presence	confront	behavior	precessity	motion	enable
redition	grant	insist	commission	assign	phenomenon
council	conduct	recentive	attorney	portion	adequate

estate	enterprise	crucial	contrary	nilent	dominate
residential	core	prominent	hesitate	dismonital	obligation
impose	profigate	motive	sustend	attain	brice
incution	dominant	glance	prodacit	promote	resolve
inspire	indication	border	version	appreciate	territory
excind	mere	controversy	anticipate	consult	conventional
tendency	massive	mascent	listure	guidance	virtually
dimension	protest	prevail	implication	jury	spite
assumption	trace	sacred	illusion	nod	rely
notion	anxious	regardless	variation	inevitable	vague

inherit	prost	swant	ethical	profound	vivid
promotion	apparatus	mutual	induce	persist	condemn
classify	peculiar	devotion	affection	disclose	adapt
explicit	merge	voluntary	heritage	texture	evidently
ultimately	bundle	inherent	ordantly	tragic	forpel
elaborate	companion	magnitude	integrate	utilize	eternal
remonistrate	chore	dilemma	privelege	modify	neglect
dispute	generous	sewage	excess	coincide	elecism
stipel	urgent	affirm	drast	compel	jure
random	devise	proscrate	recover	ally	competent

imitation	crude	loyal	confession	detach	current
retreat	boast	mentally	restraint	intact	resentment
infinite	fury	reluctant	corruption	disposition	whirl
deplete	slum	misery	lective	ritual	intuition
relinquish	legitimate	conspiracy	strive	dislution	manifestation
mode	hastily	curse	solcate	corridor	indulge
halt	hap	dis crepancy	lithic	evolve	distinctive
glimpse	authentic	apprehension	linger	exploit	declectively
prolong	vigor	ambiguous	thermal	lomulate	defect
instinct	contemplate	obstacle	candigent	shatter	evoke

momentum	idle	prosectify	enhance	garment	magnify
accentary	stubborn	ambiguity	fane	wift	notify
fringe	recorisation	cluster	assertion	remedy	refine
soduction	heap	turmoil	rotate	compatible	innovation
grief	indicator	ironic	collaborate	domain	fuel
novelty	interaction	nominate	deviation	obscure	irrelevant
mislead	stern	punish	premise	ardacy	influential
venture	formidable	synthesis	suppress	cumulative	harass
culminate	infectious	plense	endeavor	reventive	appraisal
releity	diversion	array	unanimously	feasible	gossip