1988

Accessing A Global Language: A Comparison of G-TELP and TOEFL Performance

Paula C. Kelly EdD
University of San Diego

ACCESSING A GLOBAL LANGUAGE:

A COMPARISON OF G-TELP and TOEFL PERFORMANCE

by

Paula C. Kelly

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Education

University of San Diego

1988

Dissertation Committee

William Foster, Ed.D., Director

Johanna Hunsaker, Ph.D.

Edward Kujawa, Ph.D.

Acknowledgements

To my husband Paul and my son Gavin for their support through many years of graduate school.

To my Committee for their careful review of my dissertation, and for their individual contributions to my growth as well.

To Dr. Bill Foster for inspiring me with the poetic quality of his words and thoughts, and for expanding my understanding of the critical thinking process.

To Dr. Johanna Hunsaker for the encouragement, advice, and insights she provided throughout the program.

To Dr. Ed Kujawa for the careful manner with which he handled the research class concerns and material.

Table of Contents

	Page
Acknowledgements	iv
List of Tables	viii
List of Appendices	x
Chapter	
I. Purpose and Organization	1
Statement of the Issue	1
Introduction	1
Statement of the Problem	2
Research Questions	6
Instruments	7
Method/Sample	9
Data Analysis and Results	10
Limitations of the Study	11
Definition of Terms	12
Document Organization	13
Summary	14
II. Review of the Literature	16
English As A Global Language	16
English Language Speakers	16
English Language Instruction	17

Chapter Page

English As A Language of Wider Communication 19

The Spread of English 23

The Future of English 28

Influences on Language Learning and Test Performance 30

Cultural Bias in Language Testing 30

Effect of Attitudes Toward English on Proficiency 32

Educational Measurement 34

Norm- and Criterion-Referenced Tests 34

Development and History of G-TELP 38

Validity Studies 43

Conclusion 50

III. Research Design 51

Introduction 51

Instruments 51

Subjects 53

Data Collection Procedures 55

Testing Conditions 57

Data Recording Procedures 60

Data Analysis 64

Summary 65

Chapter Page

IV. Analysis of Data 67

Introduction 67

Relationship Between G-TELP and TOEFL Overall Scores and Subscores 67

Relationship Between Subject Characteristics and Overall G-TELP and TOEFL Scores 70

Relationship Between Subjects' Overall and Section Scores on G-TELP and TOEFL and Scores on Post B Proficiency Test 86

Summary 89

V. Conclusions and Implications 91

Introduction 91

Relationship Between G-TELP and TOEFL Overall Scores and Subscores 92

Relationship Between Subject Characteristics and Overall G-TELP and TOEFL Scores 106

Relationship Between Subjects' Overall and Section Scores on G-TELP and TOEFL and Scores on Post B Proficiency Test 116

Summary 119

Implications for Further Research 120

Conclusion 123

References 126

Appendices 134

List of Tables

Table		Page
1.	Correlation Coefficients for TOEFL and G-TELP Scores	68
2.	G-TELP Mean Scores for Males and Females	70
3.	TOEFL Mean Scores for Males and Females	71
4.	F Ratios for ANOVA's of Performance on G-TELP by Sex	72
5.	F Ratios for ANOVA's of Performance on TOEFL by Sex	73
6.	Correlation Coefficients for G-TELP Score and Age	73
7.	Mean Age by G-TELP Level	74
8.	Correlation Coefficients for TOEFL Score and Age	75
9.	Level One Mean G-TELP Scores by Native Language	76
10.	Level One Mean TOEFL Scores by Native Language	77
11.	Level Two Mean G-TELP Scores by Native Language	79
12.	Level Two Mean TOEFL Scores by Native Language	80
13.	Level Three Mean G-TELP Scores by Native Language	81
14.	Level Three Mean TOEFL Scores by Native Language	82

Table Page

15. F Ratios for ANOVA's of Native Language Group Performance on G-TELP 84

16. F Ratios for ANOVA's of Native Language Group Performance on TOEFL 85

17. Correlation Coefficients for Post B, G-TELP, and TOEFL Scores 86

18. Hypothesis Acceptance/Rejection Summary 88

19. Comparison of TOEFL and G-TELP Characteristics 93

List of Appendices

Appendix		Page
I.	G-TELP Score Report	134
II.	G-TELP Answer Sheet	136
III.	G-TELP Registration Form	138
IV.	G-TELP Examinee Roster	140
V.	Letter Sent to Students Requesting Participation in G-TELP	141
VI.	G-TELP Room Report	142
VII.	Sample Data Coding Sheet	143
VIII.	Native Language Codes	144
IX.	TOEFL Score Report	145

Abstract

The ESL (English as a Second Language) field has traditionally focused on 18-22 year old foreign students who come to the United States to learn English in order to pursue an academic degree. Those students gain English proficiency, earn an American degree, and often return home to be leaders in their home countries. The author has titled these students first wave learners.

As those newly trained leaders return home and begin to pursue their goals, they are met by a group of potential followers who do not speak English. These potential followers, or second wave learners, need to acquire English skills in order to join with their leaders in accessing professional and technical resources and communicating with other partners in the development process. The second wave is less well educated than the first, will not earn a college degree abroad, and will not go abroad to study English.

First wave learners have traditionally taken the TOEFL (Test of English as a Foreign Language), which is internationally recognized yet inappropriate for the second wave learners due to its focus on academic English. The G-TELP (General Test of English as a Foreign Language) is designed to test the real world English skills required by second wave learners.

The purpose of this dissertation was to examine the relationship between scores earned by subjects on these two English language proficiency tests, the G-TELP and the TOEFL. The G-TELP is a criterion-referenced test designed to assess examinees' real world English language proficiency, while the TOEFL is a norm-referenced test designed to assess examinees' academic English language proficiency.

A total of 281 subjects were tested at five different university-based ESL (English as a Second Language) institutes throughout the country. Subjects were given the G-TELP approximately two weeks prior to taking the TOEFL in order to determine the concurrent validity of the two tests.

A moderate positive relationship was found to exist for the relationship between overall G-TELP and TOEFL scores, and for the listening and vocabulary/reading section scores as well. The grammar section score correlation coefficients indicated only a low positive correlation. Subject characteristics' effect on test scores was examined as well. No significant differences were found to exist between overall scores earned by males and females on the G-TELP and TOEFL; between subjects' age and overall scores; and among scores earned by various native language groups.

The findings of the study suggest that while there is significant overlap in the English language knowledge and skills

tested by the G-TELP and TOEFL, the tests do assess different types of English proficiency. It is hoped that the information gained through this study will support the adoption of G-TELP abroad for testing a new group of English language learners in Third World countries who are essential to the economic, social and political development of the region.

Chapter I

Purpose and Organization

Statement of the Issue

Introduction

English has become the first truly global language. It is estimated that there are more than a billion speakers of English in the world today; at least a quarter of the world's population! English at this point in history is more widely spoken and written than any other language has ever been.

The chief editor of the Oxford English Dictionary has stated that "any literate, educated person on the face of the globe is deprived if he does not know English" (McCrum, Cran, & Mac Neil, 1986, p. 39). However, the importance of English for communication purposes has spread beyond the educated elite sphere.

> It is a native language or mother tongue to millions; an official language in numerous countries where it is by no means commonly spoken by the population at large; a second language promoted by school systems in many more countries; and a lingua franca in every quarter of the globe. (Starr, 1978, p. 27)

It is clear that English now fulfills the role foreseen for Esperanto and other artificial languages as early as the seventeenth century.

English satisfies a wide variety of demands for its increasingly diverse users throughout the world. In order to keep pace with modernization and technological change, employees in all types of business throughout the world are learning English. The ability to speak English is now considered a highly desirable job skill, as seen in job advertisements from Tokyo to Paris.

Statement of the Problem

A knowledge of English provides considerable advantage to the quarter of the world's population which possesses that knowledge. However, the remaining three quarters of the world's population which does not possess a knowledge of English is at a definite disadvantage. English has become the communication medium of the First World in matters such as trade, finance, and technology. It is therefore essential to be able to use English in order to participate in international business dealings and the transfer of technology between nations.

The existing disparity between economically advantaged nations and underdeveloped nations is increased and perpetuated by the lack of English language knowledge in disadvantaged nations. This disparity raises issues of fairness and equity which are directly relevant to leadership. Leaders act as change agents, and the existing disparities in English language knowledge and the benefits it provides point to the need for change. The English as a Second Language (ESL) field must focus in some new

directions in order to reduce the disparities in the level of English language proficiency present in nations throughout the world, and this study will further that redirection of effort.

At present, the tradition of teaching 18-22 year old foreign students who come to the United States to learn English in order to pursue an academic degree is well established. Those students gain some English proficiency, and earn an American degree, which equips them to return home and become leaders in their home countries. I refer to this group of potential leaders as the first wave of English language learners.

As those newly trained leaders return home and begin to pursue their goals, they immediately encounter some difficulties. The environment to which they return consists largely of potential followers who cannot speak English, and are therefore denied access to many of the professional and technical resources necessary to gain the knowledge to meet their goals. A common body of knowledge, which often requires English language proficiency to acquire, needs to be shared by leaders and followers to progress toward the desired goals. The potential followers are the second wave of English language learners. The second wave is less well educated than the first, will not earn a college degree abroad, and will not go abroad to study English.

The second wave of language learners is a significant

element in the developmental process occurring in many Third World countries today. Until they learn English to enable them to acquire the professional and technical knowledge necessary to join with their leaders in pursuing their goals, progress toward those goals will continue to be slow.

The G-TELP (General Test of English Language Proficiency) is the first standardized American criterion-referenced test of English language proficiency designed to test that second wave of learners abroad. Its focus on real world as opposed to academic English is appropriate for the second wave whose academic preparation is less than that of the first wave. First wave learners have traditionally taken the TOEFL (Test of English as a Foreign Language), which is now internationally recognized, yet is inappropriate for the second wave of nonacademic English language learners. Reliance on the TOEFL as a testing mechanism for both first and second wave language learners has perpetuated the inequities of educational tracking systems and limited the upward mobility of the educationally disadvantaged second wave language learners.

The availability and use of the G-TELP for testing second wave English learners abroad will eventually increase their access to English language instruction and to the benefits of that knowledge. The G-TELP's prominence as a non-academic internationally recognized test of real world English language

proficiency will provide an appropriate way for their competence to be tested and recognized for the first time. The social and economic mobility of the second wave learners will be facilitated by the recognition and validation of their English language skills using the G-TELP, and they will be better prepared to assist in their nations' process of development.

This study explored the relationship between subjects' performance on the G-TELP and TOEFL. Comparison of G-TELP and TOEFL test results is an essential first step in establishing G-TELP's credibility, since TOEFL is the premier internationally recognized English proficiency test against which all new tests are compared prior to acceptance abroad. The information gained about the relationship of the two test scores and the effect of student characteristics on test performance is the first step in assisting second wave learners abroad to improve their social and economic mobility and overcome some of the inequities which currently exist.

The primary objective of this study was to measure the concurrent validity of the G-TELP and the TOEFL. The first step was to measure the English language proficiency of subjects using the G-TELP and TOEFL achievement tests and to analyze the relationship between the overall scores on each test. In addition, the relationship of G-TELP scores to several subject characteristic independent variables was analyzed. The study

also includes analysis of the relationship between scores achieved on the ALI proprietary English test and scores attained on the G-TELP.

Research Questions

This study explored the following questions:

1. What is the relationship between the subjects' overall scores and section scores on the G-TELP and their overall score and section scores on the TOEFL?

2. What is the relationship between subjects' characteristics and overall score on the G-TELP and TOEFL?

3. What is the relationship between the subjects' overall scores and section scores on the G-TELP and scores on the American Language Institute (ALI) proprietary English proficiency test?

Hypotheses

The following null hypotheses were tested in this study:

1. There is no significant relationship between subjects' overall scores and section scores on the G-TELP and their overall scores and section scores on the TOEFL.

2. There is no significant relationship between subjects' characteristics and overall scores on the G-TELP and TOEFL.

3. There is no significant relationship between subjects' overall scores and section scores on the G-TELP and scores on the proprietary ALI English proficiency test.

Instruments

The TOEFL is a product of ETS (Education Testing Service) and has been used to test the English language proficiency of non-native speakers since 1964. It was developed to measure the English proficiency of international students intending to study at colleges and universities in the United States. The TOEFL is recommended for students at the eleventh grade level or above, and is, therefore, considered too difficult for younger students.

The TOEFL consists of three sections: Listening Comprehension, Structure and Written Expression, and Vocabulary and Reading Comprehension. The test takes approximately two hours to complete. Respondents receive scaled scores for each of the three sections as well as a total score.

The G-TELP was developed by the National Education Corporation (NEC) with three goals in mind. They are: To measure a wide range of proficiencies; to focus on real-world language tasks; and to provide a score report which lists subject strengths and weaknesses. Test takers are judged by how well they fulfill certain tasks or criterion, not by comparison to each other. G-TELP also consists of three sections which parallel those of the TOEFL: Listening, Reading and Vocabulary, and Grammar. There are three proficiency levels of the test which are described as follows:

Level 1 - Authentic English in Complex Communication

Level 2 - Modified and Authentic English in Simple Communication

Level 3 - EFL Classroom English in Simple Communication

Test takers are administered the level of the test which is most appropriate for their language ability. The score report details the respondent's strengths and weaknesses referenced to specific language functions (see Appendix I). This information used in conjunction with the specific level descriptors indicates what the test taker is able to do functionally in English.

The criterion for the G-TELP is mastery of a particular skill area, which is attained if the examinee scores 75% or more on a section of the test. Respondents receive a Mastery score for the proficiency level (One, Two, or Three) as a whole as well, which must be 75% or more in each of the skill areas in order to demonstrate mastery.

The amount of time required to take the test and the number of questions on the test vary according to proficiency level. The test takes between 90 and 110 minutes to complete and consists of between 70 and 90 questions.

In addition to the scores for each test, appropriate demographic information will be requested from subjects in this study. This information will include age, sex, and native language.

Method/Sample

The TOEFL was administered to individuals enrolled in English language classes at the American Language Institute (ALI) at San Diego State University. ESL students were also included in the sample from other universities including University of Delaware, Georgetown, University of Southern Florida, and California State University, Los Angeles. The test was administered repeatedly until a sample of 55 students was accumulated for Level Three, 112 students for Level Two, and 114 students for Level One. Approximately two weeks prior to the TOEFL test administration, the G-TELP was administered to the same group of subjects. A total of 281 subjects was tested on both the TOEFL and G-TELP.

The TOEFL was taken by students at the San Diego State University Test Office and at testing centers at the other universities. Official score reports were processed by ETS and sent to ALI and the other university ESL institutes. The G-TELP was administered at ALI to students by ALI teachers who have been trained to administer the test, and by an ALI administrator at the off-campus locations. Students filled out a machine readable answer sheet which was processed by an IBM PC equipped with software which produces a test score report for each respondent (see Appendix I).

In order to encourage students to volunteer to take the

G-TELP, ALI classes were visited to request participation. Students were told that taking the G-TELP would help them practice their test taking and English language skills. A fee of $40 is normally charged for the test, but students were offered the test at no charge. Each student who volunteered to take the test filled out a registration slip in class.

At the time of test administration, subjects filled out the G-TELP answer sheet (Appendix II) and a registration form (see Appendix III) which includes demographic information about them.

It was necessary to use volunteers in this study because students should not be mandated to take an additional English proficiency achievement test. It was appropriate because demands on the subject were significant.

Data Analysis and Results

The data were collected from the G-TELP examinee roster (see Appendix IV) and from the registration form (Appendix III).

The correlational method was used to explore the relationship between the scores earned by subjects on the TOEFL and G-TELP tests. This method was appropriate for the research question because no previous research has been done on the concurrent validity of the tests. Correlation coefficients were calculated for subjects' overall scores and three section scores on both tests as well. In addition, correlation coefficients were

calculated for subjects' scores on the ALI proprietary English test and their scores on the G-TELP.

Limitations of the Study

The results of this concurrent validity study done for the particular forms of the G-TELP presently in use are valid only for those forms of the test. Additional concurrent validity studies will need to be done on succeeding forms of the test in order to generalize the findings to all administrations of the TOEFL and G-TELP done within a certain period of time. The purpose of this study is to begin to develop a knowledge base on the relationship of the test scores to each other.

The fact that subjects are volunteers represents a limitation of this study. However, the characteristics of volunteer samples discovered through studies in the United States such as level of need for social approval and level of need for achievement may not be relevant for this sample. All subjects in this study were from foreign cultures and the characteristics of volunteers may not be the same across cultures.

In addition, the students' attitudes toward each of the tests could be a limitation of the study. The TOEFL is known to the students and acknowledged to be an important factor in their entrance into an American university. On the other hand, the G-TELP is a new test and the score received on it will not influence their admission to the school of their choice. This

could affect the students' motivation to perform equally well on both tests.

A third limitation of the study may be that the students taking the test are not the audience for whom the G-TELP was designed. It was designed for a nonacademic market, and the subjects will be academic track students. However, since it is only the academic track students who take the TOEFL, it is necessary to use them as the test group in order to provide initial concurrent validation of the G-TELP.

Definition of Terms

English language proficiency: For purposes of this study, the term is used to mean the written (as opposed to verbal) competence of an individual in using the English language.

Form: A form is a unique version of a test. It is desirable to create a new form of a standardized test at least once a year so that its security will not be compromised.

Interpretation of correlation coefficients: Moderate relationship = .50 - .70; low relationship = .30 - .50; little if any relationship = .00 - .30.

Scaled score: A score to which raw scores are converted by numerical transformation (i.e. percentile ranks or standard scores).

Structure: The term structure is used in the ESL field interchangeably with the term grammar.

LWC: Language of Wider Communication. A language which is used across national borders for purposes of communication.

Document Organization

The next chapters of this paper will be organized as follows. Chapter II will contain a review of the literature which is organized into three main categories. The major category is an in-depth documentation that English is in fact a global language. This topic includes a demonstration of the current widespread use of English for various purposes, a discussion of how and why English spreads throughout the world, and a discussion of the projected future of English as a global language.

The second category in the literature review is a discussion of the influences on language learning and test performance. It will include consideration of the difficulty in minimizing the cultural bias present in language testing, and of the relationship of one's attitude toward English and its speakers to learning the language.

The final section of the literature review will focus on educational measurement. Topics covered will include a discussion of norm- and criterion-referenced tests; the development, history, and description of the G-TELP, and results of previous validation studies conducted on the TOEFL and G-TELP.

Chapter III will provide an in-depth description of how the G-TELP and TOEFL testing was carried out and how the resulting

test data was analyzed. It will include a discussion of data collection procedures, a description of the data elements, and a description of statistical methods used to analyze the data.

Chapter IV will include a description of the findings regarding each of the three hypotheses and research questions. Data will be presented which establishes the nature of the relationship between overall G-TELP and TOEFL scores. In addition, findings regarding what the relationship is between the ALI proprietary English language proficiency test and the G-TELP will be discussed. And finally, a discussion of the relationship of subject characteristics to G-TELP and TOEFL scores will be presented.

Chapter V will include conclusions and implications of the study based on the data presented in Chapter IV. Conclusions will be drawn for each of the three hypotheses, and their implications will be discussed as well. Recommendations for further study will be made as appropriate.

Summary

This study explored the relationship between scores earned by subjects taking both the TOEFL and the G-TELP. The relationship of subject characteristics to performance on the G-TELP and TOEFL was analyzed as well. And finally, the relationship between subject performance on the G-TELP and the ALI proprietary English proficiency test was investigated.

The data gathered through this study provides a valuable addition to the existing body of knowledge regarding second wave English language learners. Additional information gained regarding the second wave learners will make it possible to better meet their needs for increased English language proficiency. This study therefore contributes to the effort to support increased English language proficiency for second wave learners in underdeveloped nations. Increased English language proficiency is an important skill for participants in the development process abroad. It both increases their ability to access the body of knowledge required for their countries to progress, and enables them to communicate with essential partners in the development process. It is hoped that this study provides data about the second wave learners and thereby assists in reducing some of the disparities which exist between developed and underdeveloped nations and their peoples.

Chapter II

Review of the Literature

The universal nature of the English language and its prominence as a worldwide medium of communication has been described in several ways. Terms used frequently to indicate the widespread use of English as a global language include the terms world language (Eastman, 1973), language of wider communication (Fishman, 1977b), auxillary language (Trifonovitch, 1978), link language (Fishman, 1977a), international language (Brumfit, 1982), global language (Quirk, 1969), and additional language (Fishman, 1977b). English today is the closest thing to a world language that has ever existed.

English As A Global Language

English Language Speakers

One of the ways to demonstrate the fact that English is a global language is to calculate the number of English speakers throughout the world. Several attempts to do so have yielded similar numbers. English speakers can be divided into native and non-native speakers. Lewis and Massad (1975) estimate that there are over 250 million speakers of English as a first language, while Gage and Ohannessian (1974) estimate the number to be 275 million, Muller (1964) estimates it to be 265 million, Traumuller (1975) suggests 314 million, and Starr (1978)

estimates 350 million. Fishman (1977b) has more thoroughly researched the topic than any other American scholar and judges 300 million to be a good estimate of first-language speakers of English.

It is more difficult to estimate the number of second-language users of English, and by extension the total number of first- and second-language users of English worldwide. Choi (1980) uses Marckwardt's estimation of between 800 and 900 million speakers, and Quirk (1984) estimates that English is presently in use by 700 million people (only half of these are estimated to be native speakers of the language). Fishman (1977b) estimates at least 300 million for the non-native speaker category as well, which indicates his total would be at least 600 million first- and second- or auxillary-language speakers. All sources agree that these categories of English speakers total approximately one quarter of the world's population.

First-language speakers of English are the second most numerous language group in the world. "Only Chinese surpasses English in the number of native speakers, but Chinese is far less standardized, its speakers far more concentrated geographically" (Starr, 1978, p. 27).

English Language Instruction

Extensive research has not uncovered any attempts to estimate the total number of people studying English abroad. Attempts have

been made, however, to estimate the number of school-age students studying English abroad. Fishman (1977b) states that "the vast majority of people learning English in the world today are learning it in the secondary school" (p. 14). Gage and Ohannessian (1974) calculated that there were approximately 45 million students studying English at the primary level and 70 million students studying English at the secondary level circa 1974.

Starr (1978) indicates that those 70 million secondary school students studying English represent 76% of the 93 million secondary school students in the non-English-speaking world. The large number of school age English language learners represents a shift toward English as the most studied foreign language in many nations abroad.

Gage and Ohannessian (1974) estimate that the 115 million students studying English worldwide can be broken down by continent (p. 14) as follows:

Asia	60,000,000
Africa	20,000,000
Western and Central Europe	15,000,000
Soviet Union	10,000,000
Western Hemisphere	10,000,000
	115,000,000

"English has replaced Russian as the most studied foreign language in China, Dutch in Indonesia, Spanish in the Philippines, and both German and French in the USSR" (Starr, 1978, p. 27). Gage and Ohannessian (1974) found that English was the foreign language which most students were taught first in 56 of 102 countries surveyed.

English is the language medium through which a great deal of tertiary level study takes place. It has been estimated that 40% of the world's foreign students study in English mother tongue countries (Fishman, 1977b). The majority of these students (80%) come from non-English mother tongue countries.

English As A Language of Wider Communication

English is being used increasingly as a Language of Wider Communication (LWC). Stewart (1962) defines a language of wider communication as "a language, other than an official one, which is used for communication across language boundaries for purposes of trade and commerce within a nation" (p. 21). That somewhat narrow definition requires expansion. LWCs are used by non-native speakers across national boundaries for a wide range of purposes not limited to trade and commerce. They have a tradition of a humanistic culture which dates back to the time of the Renaissance in the West, and even earlier in the East. In addition, state of the art research in science, technology and the arts is done in LWCs (T. Donahue, personal communication, November 18, 1987).

The use of English as an official language throughout the world is an indication of its prominence as an LWC. Fisherman and Fishman (1975) define an official language as "one which is used by the government for its own internal operations and promoted through the power of state" (pp. 497-98). Banks (1975) lists English as the official language of 21 countries, and the co-official language of 16 additional countries, as of January 1, 1975. According to these figures, English is the official or co-official language in 24% of the 152 nations of the world listed. The inhabitants in nations where English has official status number over 700 million, or approximately one-third of the world's population (Starr, 1978).

The widespread use of English in research is another measure of its status as an LWC. It has been estimated that over 50% of the world's scientific research is published in English (Starr, 1978). UNESCO (1958) estimated that an even higher percentage (62%) of all scientific writing was in English. Baldauf and Jernudd (1983) estimate the percentage of English language publications in five scientific disciplines to be as follows:

Chemistry	67%
Biology	86%
Physics	85%
Medicine	73%
Mathematics	69%

These percentages represent an 11% to 22% increase over the figures gathered in a similar study in 1965.

Swales (1985) has concluded that research in English is "largely the preserve of countries where English is either the national language or the official language, of countries with an international language of scholarship such as Japan or Russia, or of those individuals who enter into international collaborative networks" (p. 5). He also suspects that English is at present largely a vehicle for dissemination of research results within the northern hemisphere. Starr (1971) states that the postwar emergence of English as the principal language of science has been related to the magnitude of American investment in research, and foresees that as other countries expand their research capacities, more discoveries will be made by non-English speakers. This may lead to the decline of English as a major vehicle for scientific research dissemination, although the need to access international communication networks will certainly still exist.

English is also widely used as a medium of instruction in both English and non-English speaking nations. Fishman (1977b) estimates that English is the medium of instruction for 14 million primary students and 11 million secondary students worldwide. These figures include mostly Asian and African nations. A study by Noss (1965) shows that in the case of Southeast Asia, English medium schooling produces greater numbers of competent English

language users than does merely studying English as one subject in the curriculum.

The worldwide status of English can also be documented by examining the circulation figures for English language newspapers in non-English mother tongue countries. "Nearly every capital city in Asia and Africa (with the exception of former French colonies) has an English language newspaper" (Bowen, 1971, p. 1). Fishman (1977b) attempted to document the circulation of English language newspapers in non-English speaking countries. He estimated that the circulation for Africa was approximately 2.9 million papers, and for Asia (including the Middle East) it was 8.4 million papers. The figures are not totally reliable, however, because they do not include all English language newspapers and are often taken from publicity documents which may be designed to attract advertisers, etc. and therefore overstate circulation figures. Also, newspapers are often shared by many individuals abroad and circulation figures are, therefore, understated in terms of numbers of people actually reading the documents.

English is the most viable medium through which ideas and messages may be presented to a worldwide audience. Books are often published in English to attract an international readership. Starr (1978) estimates that one out of every three books published in Asia and Africa appears in English. He also estimates that 70%

of international mail, telexes and cables, are written in English. In addition to the written word, English predominates in verbal communication as well. More than 60% of the world's radio programs are in English (Barnett, 1964).

An interesting use of English as an LWC, and one which will probably increase, is in the arena of international conferences. English is used widely for international conferences which involve a majority of participants for whom English is not a native language. For instance, English was the language of the proceedings of the 1955 Bandung Conference in which 29 countries of Africa and Asia represented 1,400,000,000 people. It is important to note that English was not the native language of any of the countries represented (Bowen, 1971).

The Spread of English

The predominance of English as an LWC has been attributed to "two periods of world domination by English speaking countries: British imperialism in the nineteenth century, and the economic influence of the United States in the twentieth century" (Brumfit, 1982, p. 13). Hardin (1979) also sees the world position of English as derived from the economic and political influence exerted by English speaking nations during the past two centuries. Fishman (1977b) refers to English as a "language of former colonial rule" (p. 115) in recognition of the fact that the spread of English in many parts of the world initially was as an

instrument of cultural and ideological domination. Jesperson (1938) attributed the spread of the English language to "political ascendancy" (p. 233), which can be interpreted to mean that it was to the benefit of the dominated groups to learn the language of the "master."

It is important to note, however, that the language's strong identification with England and the United States is seen by most authorities to be declining. Starr (1978) believes that in the process of adoption as an official language in parts of Asia and Africa, "English lost its exclusively British or even American connotations, thus rendering it even more acceptable as a world language" (p. 27). Starr views English as a "neutral" language as used in Asia and Africa and even attributes to it the facilitation of "cross cultural integration" on those two continents. Bowers (1986) also points to the evolution which has occurred in the use of English, "the language has become essentially delinked from its native-speaking source and increasingly depoliticized" (p. 402). This means that the continued spread of English worldwide may be decreasingly influenced by political considerations. Kachru (1984) has termed the use of English today as rather apolitical, and views its use as having fewer political, cultural, and religious connotations than the use of any other language of wider communication.

It is interesting to examine the widespread motivation for learning English today. Languages are rarely acquired for their own sake, but rather as keys to access desirable things. Fishman (1977b) describes those things as

> entree to better positions, to useful specialized knowledge, to more effective tools, to more influential contacts (and thereby to control over human and material resources), to more desirable consumer goods, to more satisfying high culture behaviors, or merely to new and different in whatever domain. (p. 115)

Bowers (1986) identifies similar benefits to be gained from a knowledge of English. He points to a working command of English as a symbol of membership in an elitist society abroad, and as a vehicle for individuals to break out of their existing place in the social structure. Starr (1978) describes the advantages of belonging to the elite society of English speakers as well. He states that, "English is the property of the educated and cosmopolitan population; it sets these groups off from the rest of their societies and gives legitimacy to their claims to represent their country before the outside world" (p. 29).

Fishman (1977b) has focused on a trend regarding motivation for learning English which has particular relevance for this study. He sees English becoming a "bread and butter skill" (p. 219) for many people who do not plan to continue their studies

beyond high school, and likens it to shorthand, typing and bookkeeping in that respect. The importance of English for many white collar jobs results in a strong incentive for growing numbers of people to learn English who will never attend college.

Fishman (1969) draws several conclusions about how English is spreading in developing nations. The first is that diffusion occurs from the top down, from elites to populace. The second is that English is learned best when there are social forces which reinforce its acquisition outside the school system. And finally, the foreign national English-speaking elites form a powerful force in support of the spread of English. The diffusion of English into their societies is facilitated more by these insiders than it is by Western outsiders, according to Fishman.

Several attempts have been made to categorize the world's nations according to the way in which they currently use English as influenced by their individual histories. The international diffusion of English may be viewed in terms of three groups. Kachru (1984) labels these the inner circle, the outer circle, and the expanding circle. The inner circle includes those areas where English is used as a first language. Countries in this category are the United States, Great Britain, Canada, Ireland, South Africa, the Caribbean, Australia, and New Zealand. The outer circle represents the institutionalized non-native varieties of English. These areas have undergone extended

periods of colonization, and the resulting English was significantly affected by contact with users of the inner circle. This group forms a community of great size and diversity and includes countries such as Nigeria, Singapore, and India. The expanding circle includes nations which have not been under the colonial rule of English speaking nations. This rapidly growing group includes Japan, Korea, Taiwan, Indonesia, Israel, and Saudi Arabia.

Quirk (1984) also categorizes the nations which use English into three categories. English is in use by three to four hundred million people who were not brought up speaking English as their native language. Most of them live in countries requiring English for external purposes. These countries are referred to as EFL (English as a Foreign Language) countries. English is used by EFL country inhabitants in contacts with both English speaking and non-English speaking countries. EFL countries include many Asian and European nations.

A second category of countries where English is also not a native language uses English widely for internal purposes such as in administration, in broadcasting and in education. In these countries, called ESL (English as a Second Language), English is usually named in the Constitution as one of the national languages in combination with the indigenous language(s). ESL countries include many Asian and African nations such as India and

Singapore. And finally, ENL (English as a Native Language) countries are defined as those for which English is a native language. They include the United Kingdom, the United States, Australia, and South Africa.

It is important not to confuse Quirk's definitions of ESL and EFL with those commonly in use in the linguistics field. English as a Second Language (ESL) refers to the use of English in contexts where it is the primary language of communication. English as a Foreign Language (EFL) refers to the use of English where it is not the primary language of communication, as is the case when it is used abroad.

Several parallels may be drawn between Quirk and Kachru's categories of English users. Kachru's inner circle English users correspond to Quirk's ENL countries. His outer circle countries appear to correspond to Quirk's ESL countries, although so few examples are given that it is difficult to be sure a complete agreement in membership in the categories exists. And finally, Quirk's third category of EFL nations appears to include both industrialized and developing nations, while Kachru's expanding circle definition focuses on developing nations.

The Future of English

Prime Minister Nehru of India said in 1981, "All regional languages must be developed and promoted. But that does not mean that English should be discarded. To do that would amount to

closing a window on the world of technology" (quoted in Bowen, 1971, p. 2). This statement highlights a trend occurring in many developing nations today, where governments are investing significant resources in establishing indigenous languages for purposes of identification, unification, and communication within national boundaries. At the same time, however, it is acknowledged that an LWC is required to permit access to and communication with the outside world, to provide the "window" described by Nehru. The importance of English to these developing nations will increasingly reside in its usefulness as a channel to the outside world.

Both Starr (1978) and Bowen (1971) predicted that English would be abandoned as an official language in developing nations as the use of indigenous languages for internal national affairs becomes more widespread. This has in fact already happened in countries like Malaysia and Tanzania. One result of the decline of English as an official language may be a drop in the number of educational institutions which use English as a medium of instruction. In addition, Starr also projects that third world writers will begin writing in their native languages and use English less in the future. Bowen predicts that governments moving toward using indigenous languages for official purposes may curtail the broadcasting of programs in English and its use in internal governmental affairs. However, he anticipates a

strong continued reliance on English by "anyone who hopes to rise on the socioeconomic ladder or exert leadership in almost any field" (p. 4). It is clear that, while the purposes for which English is used may be in transition, it will retain its importance as an LWC due to the window it provides for developing nations to the outside world.

Influences on Language Learning and Test Performance

Cultural Bias in Language Testing

Because of the widespread use of English as a LWC, there are enormous needs for teaching and assessment. The issue of whether cultural content or bias in language testing is a positive or negative factor often arises. The purpose of the language testing often determines whether cultural bias is seen as positive or negative. If the purpose of the test is to determine how well a subject can perform in an English speaking environment interacting with native speakers, it is appropriate to test for cultural as well as linguistic knowledge. In many instances the fact that cultural bias exists in a test, and the test is therefore easier for certain language groups than others, is therefore not an issue.

It is almost impossible to separate language and culture. Some cultural content is desirable as a backdrop for good language testing. However, cultural bias in testing English as a second language may yield unreliable test results. There are various

ways in which a test may be culturally biased according to Cargill-Power (1980).

Culture-coded tests have questionable validity because some extra-linguistic information is required of the student in order to answer the questions. Interpreting test stimuli such as pictures may cause problems for some examinees. An example of this danger could be a test question with a picture stimulus which causes the student to choose from a multiple choice set of responses. If the picture is of a boy smiling, there may be differences in interpretation of the picture across cultures. Americans, for instance, would probably interpret it as a happy person, whereas to Asians the smiling face could mean embarrassment or confusion. It is important, therefore, not to bias the test in favor of certain nationalities which share the American interpretation by including in the possible responses the American culturally coded interpretation of the picture.

Another example of cultural interference in ESL testing is where the test administration itself presents the subject with an unfamiliar situation. Culture shock may occur for foreign subjects taking the TOEFL or G-TELP, since it may be their first experience with a timed test. There is no research available to determine the probable magnitude of the effect of this variable on subjects' scores.

Cultural bias can occur as well in the format chosen for the test questions. The cloze test represents an example of this problem. One type of cloze test involves a systematic deletion of words from text which the student is then asked to fill in by choosing the best response in a multiple choice format. Tests which solicit single word responses may be biased in favor of certain nationalities more than others. For instance, Thai students will do poorly on cloze tests which solicit single word responses because their language structure leads them to attempt to fill in the blanks with multiple word entries (Aller, 1972) rather than the single word entries required by the test.

It is probable that there are other cultural biases present in test formats that handicap various linguistic groups which have not yet been identified. Some research has indicated that elements of the second language which are similar to elements of the first language will be relatively easy to acquire (Lado, 1957), but little research has been done to identify and relate those elements to language testing. However, other research indicates the opposite, that items which are similar but not identical may cause more confusion than those which are very different.

Effect of Attitudes Toward English on Proficiency

Little research has been done to explore the question of what the relationship is between knowing or using English and one's

attitude toward the language, but two studies do yield some interesting results. The man considered to be the father of socio-linguistics, Joshua Fishman of Yeshiva University, reported the results of a study involving high school and university students and teachers in three countries: India, Indonesia, and Israel (Fishman, 1977b).

Fishman concluded that acquiring, using, and liking English as an additional language cannot be viewed as strictly a psycho-educational process in isolation from major societal factors. A positive correlation was found to exist between subjects' attitude toward English and their use of the language. The study found that subjects' strong positive views related to national language or to nationalism were negatively related to their acquisition and use of English.

Gardner has conducted several studies which have demonstrated that students' attitudes toward the community which speaks the second language being acquired are related to second language achievement. His studies suggest that the truly successful student is motivated to become integrated with the second language community (Gardner, 1968). In addition, he found that the process of second language acquisition involves taking on behavioral characteristics of the other language community. It seems probable that this resulting change in behavior and the reaction it provoked among peers and family members could contribute to

the differential rate of progress in English proficiency experienced by many students.

Trifonovitch (1978) takes Gardner's findings one step further and attempts to demonstrate that the attitudes communicated by native speakers of English toward people learning the language influence that process. He points to a feeling of inferiority which "exists among most of us who have learned English as a second or foreign language" (p. 4). He believes this feeling of inferiority is reinforced by the "overly condescending attitude" (p. 6) exhibited by native speakers of English as they attempt to coach the non-native speaker to assist them in improving their English. This feeling of inferiority which Trifonovitch describes and the hostility which non-native speakers like him may possess toward English speakers must surely inhibit their acquisition of and performance in English.

Educational Measurement

Norm- and Criterion-Referenced Tests

There are two major types of tests used in educational research: Norm-referenced and criterion-referenced. A norm-referenced test is designed to ascertain an examinee's status in relation to the performance of a group of other examinees who have completed the test (Popham, 1978). Criterion-referenced tests have become more popular since the 1960's for measuring achievement in a different way (Millman, 1979). A

criterion-referenced test is "a test used to ascertain an individual's status with respect to a well-defined behavioral domain" (Popham, 1978, p. 41). Criterion-referenced tests draw a "random or stratified sample of items from a very precisely defined content area or domain for which the content limits are clearly specified" (Borg & Gall, 1983, p. 288). They are more useful than norm-referenced tests in diagnosing specific deficiencies in content knowledge as revealed by test results.

The TOEFL is a norm-referenced test, while the G-TELP is a criterion-referenced test. The TOEFL produces a scaled score which can be used for comparison purposes to evaluate students' English language proficiency. G-TELP produces a scaled score expressed as a percentage which indicates the portion of items that were answered correctly. In addition, the G-TELP score report produces diagnostic information regarding the respondent's proficiency strengths and weaknesses as indicated by the test results. The ability to estimate what communicative functions a subject can actually perform based on attainment of a certain criterion is the major advantage of criterion-referenced tests as opposed to norm-referenced tests.

Cartier (1968) identifies several differences in norm- and criterion-referenced tests. I have summarized the ones which are relevant to this study and commented on their applicability to the TOEFL and G-TELP.

1. The traditional norm-referenced test is designed to produce a normal distribution of subject scores. The criterion-referenced test is not designed to produce a range of scores at all, since subjects' scores are not compared with each other. The TOEFL has produced a normal range of student scores consistently since its development.

2. A norm-referenced test usually samples subjects' knowledge of a wide range of material; it is hoped that the student knows more than he/she is tested on. A criterion-referenced test attempts to test every essential behavior within a specified domain. The TOEFL is a typical norm-referenced test in this regard. By testing a range of language items across the skills (listening, reading, and grammar), the TOEFL provides a gauge of subjects' knowledge of a wide range of English material. In the case of the G-TELP, every communicative behavior has been equated to a task, and performance of the required tasks indicates mastery of a desired behavior (see Appendix I).

3. Norm-referenced tests are usually satisfied with indirect testing. A criterion-referenced test often requires the subject to produce desired behaviors or skills. Due to the nature of the skill (English language proficiency) being tested, both the TOEFL and the G-TELP rely on indirect testing. Direct

observation of English language conversations would be too costly and difficult to grade objectively.

4. A subject can pass a norm-referenced test although he/she misses a certain number of items. In the case of a criterion-referenced test, the subject is expected to get all the items right, although that expectation is often lowered to 90%. A 500-600 score on the TOEFL, which is below a 630 perfect score, is considered to be passing in that it satisfies the English language proficiency admission requirement for foreign students at American universities. The mastery or passing level on a G-TELP level test is attained when 75% of the items in all skill areas have been answered correctly.

5. Criterion-referenced tests are considered more difficult to devise and administer. However, they are considered to provide greater reliability and validity regarding subjects' ability in the applicable skills or behaviors. The TOEFL provides only a scaled score which gives no absolute indication of students' strengths and weaknesses. In contrast, the G-TELP provides a score report (see Appendix I) which outlines subjects' performance in specific skill areas and related tasks for diagnostic purposes.

6. Since criterion-referenced tests are designed to elicit the actual behavior called for by a given educational objective, multiple choice items are rarely used. However,

multiple choice items are used frequently in both the TOEFL and the G-TELP due to the difficulty in evaluating and scoring other types of questions.

Development and History of G-TELP

It is often desirable to evaluate the English language proficiency of students, trainees, or employees. This can be done in at least three ways. Subjects can be evaluated using personal interviews, which provide immediate feedback regarding their proficiency. However, this method is costly, time consuming to administer, and may be influenced by rater bias. Subjects can also be evaluated using teacher recommendations, which provide an overall evaluation regarding proficiency. The drawbacks of this method are that the criteria by which teachers evaluate students may vary, teacher standards may vary, and their assessment of students may be biased. The third method available is the use of scores from standardized tests. While the tests may be easy and inexpensive to administer, most tests are norm-referenced and the score reports often do not provide specific information regarding the test taker's capabilities. Woodward states the concern of many ESL professionals regarding the limitations of norm-referenced tests. "To say that someone is in the top 10% of the group that took a test is not very informative if we don't know what that high scorer is capable of doing. We need a

description of the tasks that can be accomplished by examinees at different score levels" (Woodward, 1980, p. 4).

The G-TELP was developed to incorporate the strengths of the three evaluation methods and to provide additional information necessary to assess specific strengths and weaknesses in English language proficiency.

The G-TELP was reviewed in Reviews of English Language Proficiency Tests by H. S. Madsen (1987). According to Madsen,

> G-TELP is a multi-level, general purpose battery, which meets assessment needs not addressed by other ESL tests in the U.S.: (a) utilizes task focused items; (b) incorporates detailed diagnostic score reporting; and (c) facilitates evaluation at varying ability levels. G-TELP's criterion-referenced task orientation coupled with its detailed diagnostic reporting of proficiency on tasks, question types, and language subskills fills a need not met by other commercial ESL/EFL tests. (pp. 34-35)

Madsen lists several advantages of the G-TELP including its "real world" situations, the good variety of speakers utilized in the listening section, and the unique criterion-referenced descriptions of tasks and skills. He makes several points under the limitations and special considerations section. He suggests that the Profile B ratio score section of the score report which

reports scores on literal, inferential, and lexical questions would need some interpreting for most students and teachers. He also suggests that additional clarification would be useful in assisting students and teachers to determine which level to select when taking G-TELP. Finally, he mentions that, like most other British and American exams, there is some national bias in that most contexts and speakers are from the United States.

The G-TELP was developed at a time when a need for tests "which attempt to duplicate as nearly as possible a real life situation in which the proficiency is normally demonstrated" (Clark, 1975, p. 10) was being identified. As Butler (1981) states, "It follows then that language testing procedures should reflect the different contents in which people use language. A person's language skills should be assessed in light of his practical needs with the language" (p. 8). It is clear, therefore, that a test such as TOEFL is not appropriate to assess the English language proficiency of policemen in Hong Kong, hotel workers in Korea, or computer technicians in Jakarta. However, the G-TELP could be used to assess the proficiency of these types of speakers.

The G-TELP provides English language assessment at three levels of proficiency. The three levels and their descriptions are as follows:

Level 3: Classroom English in Simple Communication. The

tasks at this level are typical of communicatively-oriented EFL/ESL textbooks and classroom activities and are intended for the person who has learned English primarily in the classroom.

Level 2: Authentic and Modified English in Simple Communication. The tasks at this level are based primarily on real-world sources and are intended for the person whose classroom-based learning has been reinforced by some contact with English-speaking situations.

Level 1: Authentic English in Complex Communication. The tasks at this level are taken from authentic, real-world sources and are intended for the person who will have extensive contact with native speakers.

At each of the proficiency levels, the G-TELP consists of a listening comprehension subtest, a reading/vocabulary subtest, and for all levels except 1, a grammar subtest. These skill areas are further broken down into task/structure sections which correspond to certain items on the test and are given subscores. For instance, one form of the Level 3 test is organized as follows:

Skill Area	Task/Structure
Listening	Announcements
	Descriptions
	Personal Accounts

Skill Area	Task/Structure
Reading and Vocabulary	Announcements
	Simple Biographies
	Personal Letters
Grammar	Personal Pronouns
	Simple Present Tense
	Present Perfect Tense
	Past Progressive Tense

In addition, each skill area (listening, reading and vocabulary, and grammar) is subdivided into question information types which correspond to certain items on the test and are given subscores. The question information types are literal, inferential, and vocabulary. Literal information questions ask the examinee about information which is explicitly stated in the passage, while inferential information questions require the examinee to deduce information which is not explicitly stated but implied by the passage. The vocabulary questions ask the examinee to select synonyms for words occurring in the context of the reading passage.

The G-TELP was pilot tested in five countries (Egypt, Saudi Arabia, Mexico, Japan, and the U.S.) and administered to a total of 1910 subjects world-wide between September 1984 and August 1985 (D. Brusasco, personal communication, October 25, 1985). Information on test results is available for two of the groups

tested. The test was administered to 64 Egyptians whose mean percentage score was 37.5 for grammar and 24.9 for listening, with a combined mean of 30.7 (Dr. Moktari, personal communication, January, 1985). A more ambitious pilot testing program was conducted in Japan, where 370 high school, university, and business examinees took the test. All three groups of examinees performed highest on the grammar and reading sections of the test. The two sections of the test on which the examinees received the lowest scores were vocabulary and listening (T. Hudson, personal communication, February 20, 1986).

Validity Studies

While it is possible to find descriptions for more than a dozen types of test validity, it is generally recognized (APA, 1974) that they may be categorized into three major discrete but interdependent types: Content, criterion-related, and construct.

Content validity is the degree to which a sample of items, tasks or questions on a test is representative of a defined universe of content that the test is designed to measure. There are two types of criterion-related validity: Concurrent and predictive. The criterion-related validity of a test is determined by analyzing the relationship between a score on some other variable, called a criterion, and a score on the test under consideration. In a concurrent validity study, the criterion measure is administered within a short period of time

of the test being evaluated to assure that the subjects' achievement level has not changed markedly between the respective test administrations. Predictive validity is the degree to which performance at a certain level on a test accurately predicts the later behavior of subjects. Construct validity of a test is the extent to which it measures a hypothetical construct. Examples of constructs are psychological characteristics such as intelligence, or more specifically for the purposes of this study, proficiency in English as a second language. Content-related validity is determined by examining the content of the test, while criterion and construct-related validity are usually determined by making judgments based on statistical relationships.

TOEFL Validity Studies. Studies are conducted on an ongoing basis to determine TOEFL's validity, how well the test measures a person's proficiency in English as a second or foreign language. The content validity of the test is the responsibility of the TOEFL Committee of Examiners, which is composed of six specialists in the fields of linguistics, language testing and English as a second language. The committee determines the skills to be tested, the kinds of questions to be asked, and the appropriateness of the test in terms of subject matter and cultural content. A list of specifications is developed for items appearing in various sections of the test. The specifications identify the aspects of English that are to be

tested and describe appropriate techniques for testing them (TOEFL Test and Score Manual, 1985).

The most recent and comprehensive study of TOEFL's content validity was conducted by Duran, Canale, Renfield, Stansfield, and Liskin-Gasparro (1985) and analyzed the content validity of one form of TOEFL from several different frameworks. These frameworks include the grammatical, sociolinguistic, and discourse competencies which constitute communicative competence. One factor considered was the type of speech recognition skills required to perform well on the test. They found that in parts of all three of the TOEFL subsections, statement items contained significant numbers of idiomatic expressions, which meant that the English recognition and comprehension skills being tested included more than formal academic English. On the other hand, they found that formal academic English is used in sections of the test where passages are drawn verbatim from academic texts. Duran et al. (1985) judged that the varied sample of language present throughout the test was appropriate to test the English language competence of test takers. They concluded that successful performance on the TOEFL requires a wide and appropriate range of competencies to assure English language proficiency at the level indicated by the TOEFL score.

TOEFL concurrent validity has been tested in several previous studies using various criterion measures. Maxwell (1965)

found a .87 correlation between total scores on the TOEFL and scores on a proprietary English proficiency test used at University of California, Berkeley. A similar study conducted in 1966 at Georgetown University comparing scores on TOEFL with scores on the campus proprietary English proficiency test yielded a correlation of .89 (TOEFL Test and Score Manual, 1985). Pack (1972) demonstrated moderate positive correlations between TOEFL and Michigan test scores. Upshur (1966) discovered a .89 correlation between TOEFL and the Michigan Test of English Language Proficiency scores in a study of students at San Francisco State College.

The other criterion frequently used in TOEFL concurrent validity studies is teacher ratings of student performance. Hosley and Meredith (1979) found that a low correlation existed between class grades for students at Arizona State University's language institute and TOEFL scores; no correlation exceeded .32. In the Georgetown University study previously mentioned, a .73 correlation between TOEFL scores and teacher ratings was documented. Columbia University, New York University, and the University of Michigan compared institutional rankings of students' proficiency with TOEFL scores and found correlations of .78, .87, and .76 to exist (Dizney, 1965).

The predictive validity of the TOEFL has been found to be poor with regard to academic performance. According to ETS,

> Because the TOEFL is a measure of English language proficiency, not of academic aptitude, it is inappropriate to use TOEFL scores to predict academic performance. Predictive validity studies . . . have generally yielded positive correlations between TOEFL test scores and grade point averages; however, these correlations have usually been too low to be of any practical significance. (TOEFL Test and Score Manual, 1987-1988, p. 17)

Factors other than English language proficiency which contribute to academic success include knowledge of the subject matter, academic aptitude, study skills, cultural adaptability, and financial security.

The TOEFL's construct validity is assessed by comparing the performance of native and non-native speakers of English on the test. The construct validity of the test was found to be good by Angoff and Sharan (1970). They observed that the mean scores on TOEFL of native speakers were much higher than those of foreign students who had taken the test. A large percentage of the native speakers earned maximum or near-maximum scores on the test. A study of native speaker performance on the TOEFL was also conducted by Clark (1977). Performance of native speakers clearly exceeded that of non-native speakers in the study, as

indicated by mean scores of 134 (for natives) and 89 (for non-natives) out of 150.

The Buros (1972) review of the TOEFL mentions several concerns about the test. The Listening Comprehension section of the test is criticized for measuring skills other than listening comprehension such as ability to take notes and prior knowledge of the topic. The validity of the Structure section of the test is questioned because so much communication goes on which violates grammar rules; therefore, a knowledge of grammatical structure does not insure an ability to communicate well in English. A question is raised regarding the Vocabulary section because the words do not seem to have been selected with graded difficulty and abstractness in mind. And finally, the Writing Ability section is criticized because it focuses heavily on a knowledge of grammar, and grammatical knowledge is not a good predictor of how well a student writes. These criticisms of the TOEFL would be true of other English proficiency tests which include similar listening comprehension and structure subsections.

There is not a great deal of information available regarding the validity of the G-TELP since it is a relatively new test. However, a thorough review of NEC (National Education Corporation) files does produce some relevant data.

G-TELP Validity Studies. The G-TELP's content validity was analyzed by Dr. Mohammed Mulla, Vice Rector for Academic Affairs,

University of Petroleum and Minerals, Saudi Arabia. His concerns fall into two major categories: Cultural bias and academic skills. Dr. Mulla found the test in general to be too "U.S. oriented," since it included situations and historical figures such as Ben Franklin who were not well known in Saudi Arabia. Also, much of the vocabulary seemed to be American (as opposed to British) in origin, i.e. "gas" versus "petrol" (M. Mulla, personal communication, November 3, 1984). His second concern relates to a basic issue in the field of ESL proficiency testing. It is best expressed in an excerpt from Buros' review of the listening section of the TOEFL, ". . . one wonders what is being measured. Is it understanding of English, ability to take permitted notes, recall of details, general intelligence, or prior knowledge of the topic of the lecture?" (Chase in Buros, 1972, p. 266). In the case of Saudi students taking the G-TELP, Dr. Mulla's concern was that skimming and scanning are not taught in English textbooks in Saudi Arabia. The students' lack of those academic skills could result in reduced scores on the test due to a deficiency in academic training rather than English proficiency.

It is the purpose of this study to provide concurrent validity data for the G-TELP, and no previous research has been done in this area.

The G-TELP's construct validity was assessed by administering the Level One test to a group of 79 native speakers. The mean

score was 94% and the range of scores was 89-97% (F. Davidson, personal communication, June 21, 1985). As was the case with TOEFL construct validity studies, the native speakers' scores were observed to be much higher than the foreign respondents' scores. In addition, the entire group earned near maximum scores on the test. It is noteworthy that the sample used for this study consisted largely of ESL professionals. It is probable, therefore, that their scores would exceed those of the English speaking population at large.

Conclusion

It is clear that English has become a global language and will continue in that role for the forseeable future. English is used by an increasingly heterogenous group of people worldwide as its functions are expanded. The continuing evolution of English language usage around the world has made it essential to develop new instruments to test the English language proficiency of an expanding constituency. The G-TELP is designed to address the needs of that new constituency. Once it is validated through this study in relation to the internationally accepted TOEFL, its credibility as a valid instrument will be strengthened. This should make it possible for G-TELP to better serve the needs of the increasingly diverse English language constituency.

Chapter III

Research Design

Introduction

The primary objective of this study was to measure the concurrent validity of the G-TELP and the TOEFL. The first step was to measure the English language proficiency of subjects using the G-TELP and TOEFL achievement tests and to analyze the relationship between the overall scores on each test. The study also included data gathered regarding the relationship between scores achieved on the ALI proprietary English test and scores attained on the G-TELP. In addition, data was collected regarding several subject characteristic independent variables, including age, sex, and native language, and their relationship to G-TELP scores was analyzed.

Instruments

The TOEFL is the premier internationally recognized norm-referenced test of English language proficiency for non-native speakers who plan to attend an American university. It is published by ETS (Educational Testing Service), a non-profit organization, which does a thorough job of reliability and validity testing as new test forms are administered to insure the test's continued high quality. Approximately 345,000 people registered to take the TOEFL in more than 170 countries in 1986-87.

The G-TELP is the first standardized American criterion-referenced test of English language proficiency for non-native speakers who are not following an academic track. Its focus is on testing real world rather than academic English as does the TOEFL. Approximately 10,000 people in eight countries have taken the G-TELP during the last two years.

The ALI-B is proprietary English language proficiency test administered to ALI students. It is a multiple choice test similar in design and difficulty to the Michigan English Placement Test. The ALI-B consists of four sections which total 100 points: Listening (20), Vocabulary (30), Grammar (30), and Reading (20). It is used to place students in one of six general ALI proficiency levels, and has been equated to various TOEFL scores as follows:

ALI Class Level	ALI-B Score Range	TOEFL Equivalency
101 lower elementary	20-29	350 or less
102 upper elementary	30-39	375
103 lower intermediate	40-49	400
104 upper intermediate	50-59	425
105 lower advanced	60-69	450
106 upper advanced	70-79	475

Test reliability has been established at .88.

Subjects

Subjects were students enrolled at five university-affiliated ESL institutes throughout the United States. Students enrolled at ESL institutes have not yet attained sufficient English language proficiency to be admitted to an American university. The students take the TOEFL periodically to determine whether their English proficiency is sufficient to be admitted to the university of their choice.

Two major groups of subjects participated in this study. The first group consisted of students enrolled at the American Language Institute (ALI) at San Diego State University. There were 171 subjects in this category who took both the TOEFL and G-TELP between March and November of 1987 in four separate test administrations. The number of subjects taking both G-TELP and TOEFL in a given month is presented below by G-TELP level.

Date	G-TELP Level 3	2	1	Total
March			14	14
May	3	9	18	30
July	10	22	10	42
November	25	23	37	85
TOTAL - ALI	38	54	79	171

In addition, students at four other English as a Second Language institutes affiliated with United States universities were given both the G-TELP and TOEFL as well. Numbers of subjects at the four other institutes and their G-TELP levels are listed as follows:

Date	Place	G-TELP Level 3	2	1	Total
Nov. 1987	U of Delaware	4	8	9	21
Nov. 1987	Georgetown	13	15	18	46
Nov. 1987	U of Southern Florida		21	8	29
Nov. 1987	Calif. State Univ. Los Angeles		14		14
	TOTAL	17	58	35	110

A total of 110 students at the four institutes listed above took the two tests in November, 1987.

The combined totals for both types of subjects are listed below by G-TELP level:

	3	2	1	Total
Combined Totals	55	112	114	281

A total of 281 subjects at the five ESL institutes throughout the United States took the G-TELP between March and November, 1987.

Data Collection Procedures

In an effort to gather data about as diverse a subject population as possible, subjects were used from five ESL institutes. Dr. Robert Lado of Georgetown University, director of test development for the G-TELP, and ALI administrators were asked to recommend potential university-affiliated institutes for the study. Administrators at eight institutes were contacted from the list compiled of possible institutes. Initial contact was by a letter explaining the study, which was followed up by a phone call. Four institutes in addition to ALI agreed to participate in the study: Georgetown, University of Delaware, University of Southern Florida, and California State University, Los Angeles (CSLA).

All participating institutes were given the level description for G-TELP Levels 3, 2, and 1. All five institutes decided to use all three levels in their test administration, with the exception of CSLA which chose to use only Level 2.

Subjects were given the G-TELP within two weeks of the date they were scheduled to take the institutional TOEFL at their various institutions. Subjects were given the G-TELP on a voluntary basis. Classes at the language institutes were visited to discuss the G-TELP and request student participation in the

study. In some cases a letter was sent to students as well requesting their participation in the study (see Appendix V). Students were told that taking the G-TELP would be beneficial to them because it represented an opportunity to practice English language skills and test taking behavior. They were also told that while a fee of $40 is normally charged for the test, they would be given the test free of charge in order to encourage their participation.

Response to the request for volunteer subjects was excellent in all five locations where participation was requested. In all cases, over 80% of the potential subject pool sat for the G-TELP, and in many cases nearly 100% of the potential subjects volunteered.

Procedures for administering the G-TELP are outlined in the test administrator's manual. Teachers who administered the test were trained in seminars led by ALI administrators, one of whom was a member of the initial development team for the G-TELP. Each test site (usually a classroom) was given a test administrator's manual, the registration forms, G-TELP answer sheets, G-TELP test booklets, a cassette tape for the listening section of the test, a tape recorder, and #2 pencils.

Subjects filled out two forms: A registration form (Appendix III) and a G-TELP answer sheet (Appendix II). The data from the registration forms was transferred to a coding

sheet for input into the computer. The G-TELP answer sheets are machine readable when filled out using #2 pencils. They were scored using an IBM PC equipped with software which produces a test score report (Appendix I) for each subject. The test booklets are shredded by machine after they have been used by subjects to insure continued security of the test.

Testing Conditions

The G-TELP administration manual calls for one room monitor and one proctor to be present for every 15 subjects being tested. The room monitor fills out the room report (see Appendix VI) after the test administration and is responsible for reading test instructions to the subjects. The proctor is responsible for ensuring that examinees proceed at the proscribed pace and that each examinee works independently. Teachers were trained by ALI administrators to perform these functions. Training consisted of a brief description of the test's attributes, a review of test administration procedures, and a discussion of the concurrent validity study.

A total of approximately 50 teachers were involved formally and informally in administering the G-TELP at the five locations. At Georgetown, University of Delaware, University of South Florida, and Los Angeles State there were between five and seven instructors present during the testing process. At ALI there were at least six teachers present (two for each of

the three levels being administered) at the three administrations in May, July, and November. The March administration involved only two teachers formally and two observers, since only one level of the G-TELP was given to subjects.

At four of the five institutes, each of the three levels of the G-TELP were given in separate classrooms. Separate administration for each level was essential because the length of the three subsections is different for all three levels, and the verbal instructions given to subjects vary as well. At the University of Delaware, all three levels were given to subjects in the language laboratory, which is equipped with headphones. Test instructions were read to students using the language laboratory audio system and were received through the subjects' headsets. Subjects were therefore able to take different levels of the test concurrently with minimal distraction.

Subjects were seated in alternate seats at the ALI, Southern Florida State, and Georgetown institutes to discourage cheating. At the University of Delaware it was not possible to seat the students in alternate seats due to the small size of the language laboratory. The classroom chosen for the G-TELP administration was also small at Los Angeles State and did not permit for students to be seated in alternate seats.

A room report (see Appendix VI) is filled out after each test administration. A review of the test reports for the

eight administrations involved in the study indicates that two cases of apparent cheating were reported. These instances occurred during the administration at ALI. The students were observed looking at their neighbor's answer sheet, and were asked to stop the behavior if they wanted to complete the test. The behavior stopped in both cases after the warning.

In the test administrations at all sites, subjects were given a brief break after the listening section of the G-TELP. The length of the break varied from 2 to 5 minutes at the various locations. At the Los Angeles State location, several students left the classroom briefly.

There was some variation in the attitude of teaching staff toward the G-TELP and the testing process. Some teachers viewed the G-TELP as somewhat irrelevant for their students since the majority of the subjects were university bound students. Other teachers resented the loss of class time required for subjects to take the G-TELP, since this resulted in a somewhat shorter time period to accomplish their goals. Some teachers viewed the G-TELP as an excellent opportunity for students to practice their test taking behavior and skills, and other teachers were enthusiastic about the diagnostic information which would be provided on subjects as detailed in the score reports.

Student attitudes toward and opinions of the G-TELP varied as well. The students taking Level 1 reported it to be difficult;

students taking the other two levels were more inconsistent in their reaction in the test. Those comments ranged from "very difficult" to "I really enjoyed the test." In addition, the subjects' level of anxiety appeared to be less than when taking the TOEFL because the score received on the G-TELP would not directly influence their admission into an American university as would their TOEFL score.

Data Recording Procedures

Data from the registration forms and the test score reports was entered onto a coding sheet. In addition, ALI-B scores for some of the subjects and TOEFL scores were coded onto the Fortran sheets as well. Six sources of information were required to compile a complete set of data for each subject. An example of the coding sheets can be found in Appendix VII.

The chart on the next page lists the data item names, number of digits in each item, the space numbers where the item can be found on the Fortran coding sheets, and the source of the data item. A description of the data items is as follows:

1. Test Identification number--Each G-TELP test booklet has a seven-digit test book number stamped in the upper right hand corner of the booklet. The last three digits of the test book number were used as the test identification number and entered into numbers 2-4 on the coding sheet. The first digit indicates the G-TELP test level taken: 1, 2, or 3.

	Data Item	No. of Digits	Numbers on Form	Source
1.	Test identification	4	1-4	Test Booklet
2.	Administration date	4	5-8	Registration Form
3.	Age	2	9-10	Registration Form
4.	Sex	1	11	Registration Form
5.	Native language	3	12-14	TOEFL System
6.	Years studying English prior to ALI	2	15-16	Registration Form
7.	Hours studying English per week prior to ALI	2	17-18	Registration Form
8.	Months studying at ALI	2	19-20	Registration Form
9.	G-TELP total % correct	3	21-23	G-TELP Score Report
10.	G-TELP grammar score	3	24-26	G-TELP Score Report
11.	G-TELP listening score	3	27-29	G-TELP Score Report
12.	G-TELP reading/vocabulary score	3	30-32	G-TELP Score Report
13.	TOEFL score	3	33-35	TOEFL Score Report
14.	TOEFL listening	2	36-37	TOEFL Score Report
15.	TOEFL structure	2	38-39	TOEFL Score Report
16.	TOEFL reading/vocabulary	2	40-41	TOEFL Score Report
17.	Post B score	2	42-43	ALI Student Files
18.	ALI TOEFL class	1	44	ALI Student Files

2. Administration Date--The month and year in which the test was taken were coded into spaces 5 through 8.

3. Age--The subject's age was entered in spaces 9 and 10. The age had to be calculated for each subject by using the birthdate from the registration form.

4. Sex--Subjects' sex was coded into space 11 using 0 for male and 1 for female.

5. Native Language--The three-digit identification system used by TOEFL for native languages was used for this study and coded into spaces 12-14. A copy may be found in Appendix VIII.

6. Years Studying English--The number of years the subject studied English prior to attending ALI was coded into spaces 15 and 16.

7. Hours Studying English Per Week Prior to ALI--This information was taken from the registration form and coded into spaces 17 and 18.

8. Months Studying at ALI--The number of months the subject had studied at ALI was taken from the registration form and coded into spaces 19 and 20.

9. G-TELP Total Percent Correct--The overall percentage score for G-TELP was taken from the G-TELP score report and coded into spaces 21-23.

10-12. G-TELP Sub-Scores--The G-TELP scores for each of the test's sections: Grammar, listening, and reading-vocabulary were

taken from the G-TELP score report and coded into spaces 24-32. Three spaces were allowed for each subscore. There is no grammar section in Level 1 of the G-TELP, so there was no grammar subscore for subjects who took Level 1.

13. TOEFL Overall Score--The total TOEFL score was taken from the TOEFL score report (see sample Appendix IX) and coded into spaces 33-35.

14-16. The TOEFL scores for each of the three sections of the TOEFL (listening, structure, and reading/vocabulary) were taken from the TOEFL score report and coded into spaces 36-41. Two spaces were allowed for each of the three subscores.

17. Post B Score--The two-digit post B score was taken from ALI student files and entered into spaces 42-43.

18. ALI TOEFL Class--The ALI TOEFL class level was taken from ALI student files and entered into space 44. Levels range from 101-106, with 106 being most advanced. The last digit of the level was used for coding purposes.

The following example will illustrate the coding methodology. An Indonesian male, aged 25, took G-TELP Level 2 in March 1987. He had an overall score of 73%; 71% in grammar, 75% in listening, and 74% in reading. His test booklet number was 2010765. He had been studying English for three years at 20 hours per week prior to coming to ALI. He had been enrolled at ALI for four months. His TOEFL score was 457, listening 44, structure 41, reading 39.

His Post B score was 69 and he is placed in TOEFL level 104.

The data entries on the Fortran form would be as follows:

ID#	Test Date	Age	Sex	Lang. Code	ESL Background
2765	0387	25	0	328	03 20 04

Overall G-TELP	G	L	R	TOEFL	L	S	R	B	
073	071	075	074	457	44	41	39	69	4

Data Analysis

Data was taken from the coding sheets and entered into San Diego State University's Cyber 180-830. The data was analyzed using SPSS (Statistical Package for the Social Sciences).

Data analysis was organized into three main categories. They were as follows:

1. Analysis of the relationship between subjects' overall and section G-TELP and TOEFL scores. This analysis was accomplished by calculating the product-moment coefficient r for the score relationships, since both variables are expressed as continuous scores.

2. Analysis of the relationship between subjects' characteristics and their overall G-TELP and TOEFL scores. The product-moment coefficient r was calculated for the subject

characteristic age, which is expressed as a continuous variable.

For the variable native language, an ANOVA was calculated to determine if significant differences existed in test performance for various language groups. For the variable sex, a T test was calculated to determine whether significant differences existed.

3. Analysis of the relationship between subjects' overall score on the G-TELP and overall score on the ALI Post B proficiency test. The product-moment coefficient r was calculated for this relationship as well since the variables are expressed as continuous scores.

The data analysis will be presented in the next chapter of this paper.

Summary

The English language proficiency of 281 subjects was tested by administering the G-TELP and TOEFL to them within a two-week time frame. Subjects participating were drawn from language institutes at San Diego State University, Georgetown University, University of Delaware, University of Southern Florida, and California State University, Los Angeles. In addition to test scores for the G-TELP and TOEFL, data was gathered on subject characteristics including age, sex, and native language. And finally, data was collected for an additional measure of ALI

subjects' English language proficiency, the ALI Post B test.

Correlation coefficients were calculated for three types of relationships:

1. The relationship between subjects' overall G-TELP and TOEFL scores.

2. The relationship between subjects' characteristics and overall G-TELP and TOEFL scores.

3. The relationship between scores on another measure of subjects' English language proficiency as represented by the ALI Post B test.

Chapter IV

Analysis of Data

Introduction

The G-TELP and the TOEFL were administered within a two-week time frame to a group of 281 subjects at five ESL institutes throughout the United States. The concurrent validity of the two tests was analyzed by calculating product moment correlation coefficients for the relationship between the overall and section test scores. In addition, the relationship between subject characteristics of age, sex, and native language was evaluated by calculating product moment correlation coefficients and ANOVA's. And finally, the relationship between subject performance on the Post B ALI proprietary English proficiency test and overall G-TELP and TOEFL test scores was analyzed by calculating product moment correlation coefficients. The results of the data analysis testing the three null hypotheses presented in the previous chapter will be described in this chapter.

Relationship Between G-TELP and TOEFL

Overall Scores and Subscores

The first null hypothesis to be tested in this study is: There is no significant relationship between subjects' overall scores and subscores on the G-TELP and their overall scores and subscores on the TOEFL.

Product moment correlation coefficients were calculated for the G-TELP and TOEFL overall scores and subscores. Table 1 summarizes the four correlation coefficients for each of the three G-TELP levels. Level One shows the highest correlation

Table 1

Correlation Coefficients for

TOEFL and G-TELP Scores

	Level One (n=114)	Level Two (n=111)	Level Three (n=55)
Total Scores	.70*	.68*	.62*
Grammar Section Scores	--	.46*	.37**
Listening Section Scores	.60*	.56*	.57*
Reading/Vocabulary Section Scores	.55*	.65*	.64*

* $p < .001$

** $p < .003$

coefficient for the overall scores (.70), which indicates the strongest positive relationship between total scores for the two tests. However, the relationship between overall test scores was similar for all three levels as indicated by correlation coefficients in the .62 (Level Three) to .70 (Level One) range.

All three levels of the G-TELP showed a moderate relationship between total G-TELP and TOEFL scores. The first null hypothesis is therefore rejected, since there is a significant relationship between subjects' overall test scores as evidenced by the three correlation coefficients.

The correlation coefficients for the sections on each of the tests indicate less significant relationship in general. Level Two shows the highest positive relationship between the grammar section scores (.46). Level One shows the highest positive relationship between listening section scores (.60), while Level Two shows the highest positive relationship between reading/vocabulary section scores (.65).

The grammar section correlation coefficients are the least significant of the three section coefficients (.46 and .37), and indicate a low positive correlation between the scores on the grammar section of the two tests. The listening section correlation coefficients are the most similar of the three subscores for all three G-TELP levels, and range from .56 to .60. The reading/vocabulary section scores are more varied for the three G-TELP levels, ranging from .55 to .65. The correlation coefficients for both the listening and reading/vocabulary sections are in the moderate positive correlation range. The section correlation coefficients provide further evidence for rejecting the first null hypothesis, since they all indicate that

there is a low or moderate positive correlation between the scores on the G-TELP and TOEFL. A low or moderate correlation indicates that a significant relationship does exist.

Relationship Between Subject Characteristics and Overall G-TELP and TOEFL Scores

The second null hypothesis to be tested in this study is: There is no significant relationship between subjects' characteristics and overall scores on the G-TELP. The three subject characteristics under consideration are sex, age, and native language.

G-TELP mean scores for males and females are presented in Table 2. Females earned higher mean scores than the males for all three levels of the G-TELP. The highest mean scores for both males (72.39) and females (81.00) were earned on Level Three of the G-TELP.

Table 2

G-TELP Mean Scores for Males and Females

	Male	Female
Level One (n=124)	57.06 (n=73)	58.12 (n=51)
Level Two (n=121)	65.54 (n=70)	67.82 (n=51)
Level Three (n=62)	72.39 (n=44)	81.00 (n=18)

Table 3 lists TOEFL mean scores for males and females. Females earned higher mean scores than the males on the TOEFL for all three G-TELP level groups. The highest mean TOEFL scores were earned by the G-TELP Level One males (542.11) and females (547.75).

Table 3

TOEFL Mean Scores for Males and Females

	Male	Female
Level One (n=114)	542.11 (n=67)	547.75 (n=47)
Level Two (n=110)	488.99 (n=65)	489.89 (n=45)
Level Three (n=55)	429.46 (n=37)	432.44 (n=18)

An Analysis of Variance (ANOVA) was calculated to determine whether a significant difference exists for males and females on G-TELP overall test scores. Table 4 presents the calculated F ratios, relevant test statistics, and probability levels for the three G-TELP level groups. The statistics indicate that a significant difference was found only for Level Three, which had an F ratio of 4.115. The Level One and Two calculated F ratios did not exceed the test statistics for those levels, which

indicates that there is not a significant difference between overall scores on G-TELP earned by males and females for those levels.

Table 4

F Ratios for ANOVA's of Performance on G-TELP by Sex

	Calculated F Ratio	Test Statistic	Probability
Level One (n=124)	.228	3.92	.6338
Level Two (n=121)	1.056	3.92	.3063
Level Three (n=62)	4.115	4.00	.0469

An ANOVA was also calculated to determine whether a significant difference exists for males and females on TOEFL overall test scores. Table 5 presents the calculated F ratios, relevant test statistics, and probability levels for the three G-TELP level groups. The statistics indicate that no significant difference exists between TOEFL overall scores earned by males and females for any of the three G-TELP level groups.

Table 5

F Ratios for ANOVA's of Performance

on TOEFL by Sex

	Calculated F Ratio	Test Statistic	Probability
Level One (n=114)	.401	3.92	.5277
Level Two (n=110)	.010	3.96	.9187
Level Three (n=55)	.047	4.04	.8291

The second subject characteristic to be considered is age. Table 6 shows the correlation coefficients for the relationship

Table 6

Correlation Coefficients for G-TELP Score and Age

Level One (n=120)	-.13
Level Two (n=120)	-.07
Level Three (n=62)	-.16

between G-TELP overall score and age for each of the three G-TELP level groups. Little if any correlation is present as indicated by the slightly negative correlation coefficients for all three groups. There is little or no relationship between overall G-TELP scores and subjects' age.

Mean age was also calculated for each of the three G-TELP level groups. Table 7 presents the mean age for each of the groups. The mean age was the lowest for Level Three (21.98) and the highest for Level One (26.08).

Table 7

Mean Age by G-TELP Level

Level One (n=121)	26.08
Level Two (n=120)	23.68
Level Three (n=62)	21.98

Correlation coefficients were also calculated for the relationship between overall TOEFL score and age. Table 8 shows the coefficients for the three G-TELP levels. Little if any correlation is present as indicated by the slightly negative

correlation coefficients for all three groups. There is little or no relationship between overall TOEFL scores and subjects' age.

Table 8

Correlation Coefficients for TOEFL Score and Age

Level One (n=111)	-.12
Level Two (n=109)	-.01
Level Three (n=55)	-.19

It is important to note the small n for many of the native language groups prior to describing the results for this third subject characteristic. The third subject characteristic considered was native language. Table 9 lists mean G-TELP scores by native language for 113 Level One subjects, and the number of subjects in each native language group. Nineteen native language groups are represented by Level One subjects whose mean scores range from 76.00 (Swedish) to 48.00 (Napali).

Mean scores were also calculated for native language groups of 114 subjects in G-TELP Level One who took the TOEFL. Table 10 lists the native language group, mean TOEFL score, and number of

Table 9

Level One Mean G-TELP Scores by Native Language

Native Language	Mean	n
Swedish	76.00	2
Italian	72.00	3
German	69.45	11
Basque	66.00	1
Vietnamese	64.00	1
Spanish	61.67	6
Portuguese	57.40	5
Swahili	57.00	1
Indonesian	56.61	23
Farsi	55.00	4
Hebrew	53.33	3
Arabic	53.29	7
Thai	53.20	5
Icelandic	53.00	1
Japanese	50.97	30
Chinese	50.67	3
Turkish	50.50	2
Korean	50.00	4
Napali	48.00	1
TOTAL		113

Table 10

Level One Mean TOEFL Scores by Native Language

Native Language	Mean	n
Swedish	617.00	1
Italian	616.67	3
Icelandic	597.00	1
Swahili	590.00	1
Portuguese	580.00	5
German	577.64	11
Basque	573.00	1
Turkish	558.50	2
Spanish	556.50	6
Vietnamese	553.00	1
Hebrew	551.00	3
Korean	541.33	3
Thai	532.20	5
Indonesian	531.27	22
Arabic	525.00	6
Farsi	524.33	3
Chinese	521.33	3
Japanese	518.27	26
Napali	480.00	1
TOTAL		114

subjects in each native language group. Nineteen native language groups are represented by Level One subjects whose mean TOEFL scores range from 617.00 (Swedish) to 480.00 (Napali).

Table 11 lists mean G-TELP scores by native language group for 106 Level Two subjects, and the number of subjects in each native language group. Fourteen native language groups are represented by Level Two subjects whose mean scores range from 82.00 (Italian) to 62.09 (Arabic).

Mean scores were also calculated for native language groups of 95 subjects in G-TELP Level Two who took the TOEFL. Table 12 lists the native language group, mean TOEFL score, and number of subjects in each native language group. Fourteen native language groups are represented by Level Two subjects whose mean TOEFL scores range from 540.00 (Italian) to 466.38 (Japanese).

Table 13 lists mean G-TELP scores by native language group for 65 Level Three subjects, and the number of subjects in each native language group. Eleven native language groups are represented by Level Three subjects whose mean scores range from 513.00 (Italian) to 387.60 (Indonesian).

Table 14 lists mean TOEFL scores by native language group for 53 Level Three subjects, and the number of subjects in each native language group. Ten native language groups are represented by Level Three subjects whose mean scores range from 513.00 (Italian) to 387.60 (Indonesian).

Table 11

Level Two Mean G-TELP Scores by Native Language

Native Language	Mean	n
Italian	82.00	1
Korean	80.00	1
Greek	76.50	2
Thai	73.67	3
Turkish	71.33	3
Farsi	70.33	3
Portuguese	70.33	3
Fulani	70.00	1
Chinese	69.00	7
German	69.00	4
Indonesian	68.82	11
Spanish	67.29	7
Japanese	62.66	38
Arabic	62.09	22
TOTAL		106

Table 12

Level Two Mean TOEFL Scores by Native Language

Native Language	Mean	n
Italian	540.00	1
Portuguese	521.00	3
Spanish	516.67	6
Greek	513.00	2
Farsi	500.00	2
Thai	498.67	3
Fulani	497.00	1
Korean	497.00	1
German	493.33	3
Turkish	488.00	3
Indonesian	485.82	11
Arabic	482.28	18
Chinese	480.43	7
Japanese	466.38	34
TOTAL		95

Table 13

Level Three Mean G-TELP Scores by Native Language

Native Language	Mean	n
Italian	97.00	1
German	94.00	1
Mende	86.00	1
Farsi	84.00	1
Chinese	83.33	3
Spanish	82.64	11
Turkish	78.00	2
Japanese	77.80	15
Portuguese	69.00	2
Arabic	66.47	19
Indonesian	66.20	5
TOTAL		65

Table 14

Level Three Mean TOEFL Scores by Native Language

Native Language	Mean	n
Italian	513.00	1
German	507.00	1
Portuguese	462.00	2
Spanish	456.70	10
Mende	447.00	1
Turkish	436.50	2
Japanese	428.47	15
Arabic	418.00	13
Chinese	398.00	3
Indonesian	387.60	5
TOTAL		53

The classification systems for the diffusion of English in nations throughout the world developed by Kachru and Quirk and discussed in the literature review portion of this paper have been applied to the mean score data in Tables 9-14. It was necessary to make an inference from the subject's native language to country of origin because the country information was not available.

The outer circle group was represented by African nations in this study. Swahili, Fulani, and Mende are the only outer circle (Kachru) or ESL (Quirk) native languages in the group of subjects tested. Swahili is most likely to be spoken in Kenya and Tanzania, which are both former British colonies (A. Johns, personal communication, March 16, 1988). Swahili was classified as outer circle or ESL because the subject's probable country of origin was a British colony and English is used as one of the official languages for internal communication purposes. Similarly, Fulani is spoken in Nigeria, and Mende is most likely to be spoken in Ghana (T. Donahue, personal communication, March 8, 1988). They were classified as outer circle and ESL because Nigeria and Ghana are former British colonies and one of the countries' official languages is English.

The other native language groups are classified as expanding circle (Kachru) and EFL (Quirk) countries. Those countries have not been under the colonial rule of English speaking nations and use English for external communication purposes. The expanding

circle includes advanced EFL nations such as the many European nations (Italy), Spain, Germany) represented by the subjects, as well as many Asian nations (Korea, Indonesia, Japan, Thailand), and Middle Eastern nations (Israel and Turkey). No inner circle (Kachru) or ENL (Quirk) nations are represented in the lists.

An ANOVA was calculated to determine whether a significant difference exists for native language groups on overall G-TELP scores. Table 15 summarizes the calculated F ratio, relevant test statistic, and probability level for each of the three G-TELP

Table 15

F Ratios for ANOVA's of Native Language Group Performance on G-TELP

	Calculated F Ratio	Test Statistic	Probability
Level One (n=113)	2.603	1.70	.0015
Level Two (n=106)	1.029	1.87	.4309
Level Three (n=61)	1.786	2.04	.0878

level groups. A significant difference was found for only Level One, since the calculated F ratio exceeded the appropriate test

statistic. In the case of Level Two, the high probability level caused the difference not to be significant as well. No significant difference was found to exist for Level Two and Level Three subjects.

An ANOVA was also calculated to determine whether a significant difference exists for native language groups on overall TOEFL scores. Table 16 presents the calculated F ratio, relevant test statistic, and probability level for each of the three G-TELP level groups. A significant difference was again found to exist for only Level One. No significant difference was found to exist

Table 16

F Ratios for ANOVA's of Native Language Group Performance on TOEFL

	Calculated F Ratio	Test Statistic	Probability
Level One (n=104)	2.424	1.70	.0035
Level Two (n=95)	1.230	1.90	.2740
Level Three (n=54)	1.853	2.08	.0796

for Levels Two and Three subjects. The high Level Two probability level caused the difference to be insignificant as well.

Relationship Between Subjects' Overall and Section Scores on G-TELP and TOEFL and Scores on Post B Proficiency Test

The third null hypothesis to be tested in this study is: There is no significant relationship between subjects' overall and section scores on the G-TELP and scores on the proprietary ALI English proficiency test entitled "Post B."

Product moment correlation coefficients were calculated for Post B, G-TELP, and TOEFL scores. Table 17 presents the correlation coefficients for the three G-TELP level groups for

Table 17

Correlation Coefficients for Post B, G-TELP, and TOEFL Scores

	G-TELP Scores	TOEFL Scores
Level One	-.05 **** (n=37)	.19 *** (n=34)
Level Two	.56 * (n=32)	.59 * (n=30)
Level Three	.59 ** (n=21)	.67 * (n=20)

* $p < .001$
** $p < .003$
*** $p < .139$
**** $p < .380$

the Post B and G-TELP or TOEFL tests. The correlation coefficients indicate there is little if any correlation or relationship between the Post B and G-TELP or TOEFL overall scores for Level One subjects. For Level Two and Three there is a moderate positive correlation between both the G-TELP and TOEFL overall scores and the Post B scores. The null hypothesis is therefore accepted for Level One subjects and rejected for Level Two and Three subjects.

The highest positive correlation was found to exist for Level Three subject Post B and TOEFL scores. The correlation coefficients for each of the G-TELP level groups are in similar ranges for both G-TELP and TOEFL, indicating the correlation of Post B test scores with each of the other overall test scores was of similar magnitude. For instance, the correlation coefficients for Level Two Post B and G-TELP and TOEFL scores are .56 and .59 respectively, which indicates that a moderate positive correlation exists for Post B and the other two overall test scores for the Level Two group subjects.

The means were also calculated by G-TELP level subject responses to the three questions. The mean number of years English had been studied prior to attending ALI was largest for the Level One group (5.72) and smallest for the Level Three (4.28) group. The mean number of hours per week English had been studied prior to attending ALI was the largest for the Level One group (13.69) and smallest for the Level Three group (6.71). And finally, the

mean number of months subjects have studied at ALI is largest for Level One subjects (4.03) and smallest for Level Three subjects (2.48).

Table 18 presents a summary of the hypotheses and their acceptance or rejection status.

Table 18

Hypothesis Acceptance/Rejection Summary

First Null Hypothesis:	No relationship between overall and subsection G-TELP and TOEFL scores.	
Overall scores	-	Reject the null
Listening scores	-	Reject the null
Reading/Vocabulary scores	-	Reject the null
Grammar scores	-	Reject the null
Second Null Hypothesis:	No relationship between subject characteristics and overall G-TELP and TOEFL scores.	
Overall G-TELP and Sex	-	Accept for Levels One and Two Reject for Level Three
Overall TOEFL and Sex	-	Accept for all three levels
Overall G-TELP and Age	-	Accept for all three levels
Overall TOEFL and Age	-	Accept for all three levels
Overall G-TELP and Native Language	-	Reject for Level One Accept for Levels Two and Three
Overall TOEFL and Native Language	-	Reject for Level One Accept for Levels Two and Three
Third Null Hypothesis:	No relationship between overall G-TELP scores and Post B scores	
Accept for Level One Reject for Levels Two and Three		

Summary

Data was collected for a group of 281 subjects who took the TOEFL and G-TELP within a two-week time frame. The data was analyzed to test three hypotheses regarding the relationship between G-TELP test scores, subject characteristics, and another test of English language proficiency.

The first null hypothesis was rejected. The correlation coefficients indicated that there is a moderate positive correlation between subjects' overall G-TELP and TOEFL scores. In addition, a moderate positive correlation was found to exist for the listening and reading/vocabulary section scores as well. The grammar section scores yielded a low positive correlation. The hypothesis that there is a relationship between subjects' overall and subsection G-TELP and TOEFL scores was accepted.

The second null hypothesis was accepted for some of the population and rejected for other parts of it. No significant difference was found to exist between overall G-TELP scores earned by males and females for Levels One and Two. However, a significant difference was found to exist for overall G-TELP scores earned by males and females who took Level Three.

With regard to overall TOEFL scores, no significant difference was found to exist between scores earned by males and females for any of the three G-TELP level groups.

Regarding subjects' age and overall G-TELP scores, little or no relationship was found to exist between subjects' age and G-TELP score for all three G-TELP level groups. The null hypothesis that no significant difference exists between G-TELP scores earned by different age groups was accepted in this instance.

With regard to subjects' native language groups and overall G-TELP scores, a significant difference was found to exist in G-TELP scores for different native language groups only for Level One. No significant difference was found to exist for Level Two and Three subjects. The null hypothesis is therefore accepted for Levels Two and Three and rejected for Level One.

Regarding subjects' native language groups and overall TOEFL scores, a significant difference was found to exist in TOEFL scores for different native language groups only for Level One. No significant difference was found to exist for Level Two and Three subjects. The null hypothesis was again accepted for Levels Two and Three and rejected for Level One.

The third null hypothesis to be tested is that there is no significant relationship between subjects' overall G-TELP scores and Post B scores. The null hypothesis was accepted for Level One subjects, and rejected for Levels Two and Three.

Chapter V

Conclusions and Implications

Introduction

The concurrent validity of the G-TELP and TOEFL was analyzed by calculating product moment correlation coefficients for the relationship between overall and section test scores. A moderate positive relationship was found to exist for overall scores, and for Listening and Reading/Vocabulary section scores. A low positive correlation was found to exist for Grammar section scores.

The relationship between subject characteristics of sex, age, and native language and overall G-TELP and TOEFL scores was explored as well. No significant difference was found between overall G-TELP scores earned by males and females for Levels One and Two, while a significant difference was found for Level Three. In the case of total TOEFL scores, no significant difference was found between scores earned by males and females for any of the three G-TELP levels. With regard to subjects' age, little if any correlation was found to exist for all three G-TELP levels between age and total G-TELP and TOEFL scores. And finally, no significant difference was found between total G-TELP scores earned by various native language groups for Levels Two and Three, while a significant difference was found for Level One. In the

case of total TOEFL scores for various native language groups, a significant difference was again found to exist for Level One only.

The relationship between Post B and G-TELP and TOEFL overall scores was also explored. A moderate positive correlation between both G-TELP and TOEFL total scores and Post B was discovered for Level Two and Three. Little if any correlation was found to exist for Level One scores.

The following discussion provides interpretation of these findings and concludes by suggesting areas for further research.

Relationship Between G-TELP and TOEFL

Overall Scores and Subscores

The first null hypothesis to be tested in this study was: There is no significant relationship between subjects' overall scores and subscores on the G-TELP and their overall scores and subscores on the TOEFL. This null hypothesis is rejected due to the fact that the product moment correlation coefficients calculated for the overall scores and subscores for the two tests indicate that a significant relationship does exist between the scores.

In order to interpret the section correlation coefficients, it is useful to examine the format of each of the G-TELP and TOEFL sections. A summary comparison of TOEFL and G-TELP characteristics is found in Table 19. The number of questions, percent of total test questions this number represents, and time

required to complete the questions is presented for each section and the total tests.

Table 19

Comparison of TOEFL and G-TELP Characteristics

	TOEFL	G-TELP LEVELS One	Two	Three
Grammar Section				
Number of questions	40	--	15	16
Percent of total test questions	27	--	18	23
Time allowed to complete (in minutes)	40	--	20	20
Listening Section				
Number of questions	50	27	30	22
Percent of total test questions	33	30	35	31
Time allowed to complete (in minutes)	45	30	40	25
Vocabulary and Reading Section				
Number of questions	60	63	40	32
Percent of total test questions	40	70	47	46
Time allowed to complete (in minutes)	60	75	50	45
Total Test				
Number of questions	150	90	85	70
Time allowed to complete (in minutes)	145	105	110	90

Grammar sections. The Grammar section coefficients for the two G-TELP levels are .46 for Level Two and .37 for Level Three. The Grammar section coefficients are the lowest of the three section coefficients, and indicate a low positive correlation between the scores on the Grammar section of the two tests. The Grammar sections of the two tests will be analyzed to assist in explaining these findings.

The Grammar section of the G-TELP consists of 15 or 16 (depending on the level of the test) sentence completion questions. Subjects are asked to choose the best one- or two-word answer to complete the sentence from four alternatives. This section requires 20 minutes to complete.

The Grammar section of the TOEFL is broken into two parts. The first part consists of 15 cloze questions which are incomplete sentences. Subjects are asked to choose the correct word or words from four alternatives which best completes the sentence. The second part of the Grammar section includes 25 sentences which have four underlined words or phrases. Subjects are asked to identify the one underlined word or phrase that must be changed in order for the sentence to be correct. This section of the test requires 40 minutes to complete.

In summary, both the TOEFL and G-TELP Grammar sections include approximately 15 sentence completion questions. As Table 19 indicates, the number of questions in this section

represent approximately the same percentage of the total questions for each test: 27% for TOEFL, and 18% to 23% for G-TELP. The TOEFL includes an additional 25 questions and 20 minutes more are allowed to complete this section of the test than is the case for G-TELP. This means that over 50% of the type of questions found in the TOEFL Grammar section are not found in the G-TELP Grammar section. The fact that there is this significant difference in the format and composition of the test may help explain why the correlation coefficients for this section are low.

Subjects are required to answer an additional type of question in the TOEFL Grammar section which is not present in the G-TELP Grammar section. This probably leads to differential performance by subjects on the Grammar section of the two tests, since all subjects will not do equally well on the additional section present in the TOEFL. This differential performance on the two Grammar sections due to differences in section composition would result in low correlation coefficients for the Grammar section.

Another reason that the correlation coefficients are low for the Grammar sections is the difference in the range of grammar items being tested in each test. In the case of the G-TELP, a somewhat narrow subset of grammar is tested in order to provide the desired diagnostic information. Specific grammar tasks were chosen to represent a certain skill area in each section of the

test, which results in a somewhat restricted range of grammar items being tested. The TOEFL tests a wider range of grammar items than does the G-TELP, which could contribute to the low correlation between the Grammar sections of the two tests.

A third factor which could contribute to the low Grammar section correlation coefficients is a difference in the level of grammar points tested in the two sections. The TOEFL may assess a higher level of grammar points than does the G-TELP because it is targeted at a higher level audience than is the G-TELP. This line of reasoning is supported by the fact that the coefficient for Level Two (.46) is higher than the coefficient for Level Three (.37). Level Two, the higher level G-TELP test, is more positively correlated with TOEFL than the Level Three test, as would be expected if TOEFL assesses the type of higher level skills more likely to be found in the Level Two test. A difference in the level of difficulty of grammar points tested would lead to dissimilar performance on the two tests by subjects, which would in turn affect the correlation coefficients. A thorough content analysis of item level for both tests would be required to determine whether this is an accurate supposition.

Listening sections. The next section of the tests to be analyzed is the Listening section. The Listening section score correlation coefficients indicate a moderate positive relationship between scores achieved on the two sections, and are similar for

all three levels: .60 for Level One, .56 for Level Two, and .57 for Level Three. There are several factors which explain why the correlation coefficients were not higher. The first factor involves a difference in format of the G-TELP and TOEFL Listening sections.

The Listening section of the TOEFL is composed of three sections: A, B, and C. Part A requires subjects to listen to short sentences and then answer multiple choice questions about the sentences. Part B asks subjects to listen to a short conversation between two speakers, and then to choose the response to a question regarding the conversation asked by a third speaker from four possible alternatives. Part C requires subjects to listen to short talks and conversations and to respond to four-part multiple choice questions about the talks and conversations. Part A includes 20 questions, Part B consists of 15 questions, and Part C includes 15 questions. Examinees are allowed 45 minutes to complete the Listening section of the TOEFL.

The Listening section of the G-TELP is constructed somewhat differently. It consists of between five and seven spoken passages, depending on the level of the test. There are between four and seven questions which are asked about each of the passages which are read to the subjects. The questions are read to the subjects prior to the reading of the passage. Each of the questions about the passages is a four-part multiple choice

question. This section consists of 22-30 questions in total, depending on the level of the test, and requires between 25 and 40 minutes to complete, again depending on the level of the test.

In summary, the Listening section of the TOEFL (50 questions) is longer than the same section of the G-TELP (22-30 questions), and examinees are given longer to complete the TOEFL Listening section (45 minutes) than they are to complete the same section of the G-TELP (25 to 40 minutes). The number of questions in this section represents approximately the same percentage of the total questions for each test: 33% for TOEFL, and between 31% and 35% for G-TELP.

The question item types in this section are very different for the two tests. In the G-TELP, subjects hear the questions being asked about the passages prior to reading the passages; this allows them to know what information they are listening for in the passages. In the case of the TOEFL, the questions about the passage are read to subjects after they hear the passages. This difference in test format could lead to differential performance on the two tests by subjects, and thereby result in only a moderate relationship between scores earned by subjects on the two tests.

Another factor which could lead to differential subject performance on the Listening sections of the two tests is the length of the passages and the number of questions asked per

passage on each test. The passages read to subjects are longer in the G-TELP than they are in the TOEFL, which requires subjects to concentrate for a longer period of time on material in the same passage. The G-TELP Listening section includes fewer passages than does the TOEFL, and more questions are asked per passage in the G-TELP as compared to the TOEFL. In this way the content is more concentrated in a few passages in the G-TELP, and it is more widely distributed in a greater number of shorter passages in the TOEFL. This difference in composition of the two sections could lead to differential performance by subjects, and thereby result in only moderate correlations between scores earned on the Listening sections of the two tests.

An additional factor which may have contributed to differences in performance by subjects on the two sections is the use of pictures in the sections. Pictures are used to set the scene in all three levels of the G-TELP; they are not used at all in the TOEFL. Cargill-Powers (1980) has stated there may be differences in interpretation of pictures used in language tests across cultures which may affect test scores. The difference in the use of pictures in the two tests could contribute to differential levels of performance by subjects and thereby influence the magnitude of the correlation coefficients.

Vocabulary and Reading sections. The Vocabulary and Reading section of the two tests is the last section to be analyzed. The Vocabulary and Reading section scores were found to have a moderate positive relationship for all three G-TELP levels as follows: .55 (Level One), .65 (Level Two), and .64 (Level Three). As was the case with the other two sections of the tests, there are differences in the composition of the two tests which may have influenced the correlation coefficients.

The Vocabulary and Reading Comprehension section of the TOEFL is composed of two parts. The 30 questions in the first part consist of sentences which include an underlined word or phrase. Subjects are asked to choose from four alternatives the one word or phrase which best retains the meaning of the original sentence if it is substituted for the underlined word or phrase. The second part of this section consists of 30 questions which are based on several short passages. Examinees are asked to respond to four-part multiple choice questions about the passage. The Vocabulary and Reading section of the TOEFL requires 60 minutes to complete.

The Vocabulary and Reading section of the G-TELP consists of 4 or 7 passages, depending on the level of the test being examined. The passages are followed by comprehension and vocabulary questions, and subjects are asked to choose the best answer from among four choices. The section includes between 32

and 63 questions, and requires between 45 and 75 minutes to complete, depending upon the level of the test under consideration.

The Vocabulary and Reading section of the TOEFL is shorter (60 questions) than Level One of the G-TELP (63 questions), and longer than Level Two (40 questions) and Three (32 questions) of the G-TELP. This section represents approximately the same percentage of total test questions for TOEFL (40%) and Levels Two (47%) and Three (46%) of G-TELP. The Vocabulary and Reading section of the Level One G-TELP test accounts for 70% of the total test questions.

The Vocabulary and Reading section of the G-TELP Level One is longer in terms of numbers of questions asked and length of time allowed to complete the test and requires subjects to process more reading material in order to complete the test. This may help account for the lower correlation coefficient of .55 found for Level One as compared to .65 and .64 for Levels Two and Three respectively.

The format of the two Vocabulary and Reading sections is different for G-TELP and TOEFL as described in the preceding paragraphs. The word substitution questions found in the first part of this TOEFL section are not present in G-TELP. This difference in section format could lead to differential performance by subjects on the two tests, which in turn could

result in only moderate relationships between scores earned on the two sections as indicated by the correlation coefficients.

It is important to note that for all three sections and the total score correlations, the n size of 55 for Level Three is approximately half the n size of 114 and 111 for Levels One and Two. This difference in n size could have affected the correlation coefficients as well.

Total score correlations. The correlation coefficients for the total scores earned by subjects on the TOEFL and G-TELP are as follows: .70 for Level One, .68 for Level Two, and .62 for Level Three. These coefficients indicate moderate positive relationships between the total scores.

Since the G-TELP is a unique test in that it is a criterion-referenced English proficiency test which provides diagnostic information, it is not possible to compare the correlation coefficients calculated for TOEFL and G-TELP with other concurrent validity studies involving TOEFL and similar tests. The information available regarding studies conducted involving TOEFL and ESL institutes' proprietary English language proficiency tests indicates the correlations were in the .87 to .89 range (TOEFL Test and Score Manual, 1985). The tests involved in these studies were designed to test the same audience as the TOEFL tests, and are constructed similarly to the TOEFL for that reason. It is therefore not surprising that

the correlation coefficients calculated for TOEFL and G-TELP were lower than those found in studies involving TOEFL and other similarly targeted and constructed tests.

The G-TELP and TOEFL are designed to test different populations and types of English language proficiency, so very high correlations would not be expected. However, the author would have predicted somewhat higher correlations at approximately the .75 range for the TOEFL and G-TELP. The .75 range would have indicated the correlation for the two tests was less than that calculated for TOEFL and similar tests, but higher than the coefficients actually calculated in this study which fall in the .62 (Level Three) to .70 (Level One) range. There are several factors which may have contributed to the correlations being less strong than predicted.

The first factor is the placement of subjects into correct G-TELP levels. The subjects were placed in G-TELP levels by personnel at the five institutes based on the level descriptors for beginning, intermediate and advanced or Levels 3, 2, and 1. In general, students were placed at a level higher than the appropriate one based on their proficiency level. Increased experience with the test has indicated that the descriptors need to be revised downward somewhat, and a new Level Four is being developed which is classified as a true beginning level. Some of the Level One material is extremely close to a native speaker

proficiency level, and few foreign students or other non-native speakers would ever attain mastery at that level. Therefore, since students were often not placed in correct levels initially, their scores were probably not as high as they would have been for the correct level test, and the resulting correlation coefficients were probably suppressed as well. It is probable that the total score correlation coefficients would have been higher and closer to the author's predicted level if the subjects had been placed more accurately in the correct G-TELP level.

Another factor which may have influenced subjects' total scores on G-TELP and TOEFL and the resulting coefficients is the extent to which bias exists within either or both test(s) and causes certain cultural or language groups to perform better on the test than others. ETS was contacted for information regarding prior studies which may have focused on the issue of cultural or linguistic bias. According to the TOEFL Associate Program Director,

> As you know, as a test of English skills, the TOEFL may differentially impact some populations whose languages are related to English. However, the term bias implies differentially affecting populations of comparable skills. To my knowledge no studies of this exist. Moreover, as a test of language skills, the term bias may not be

> appropriate for language-related differences. My understanding of research in this area is that it generally deals with differences related to variables that should not be related to the construct under measure. So, too, as culture and language are related, it would be difficult to study bias in this area. (G. E. De Mauro, personal communication, February 17, 1988)

De Mauro's comment regarding the difficulty involved in studying cultural bias in language testing when language and culture are so closely intertwined may explain the lack of research in the field. Most authorities agree that cultural bias exists in the process of language testing, but the nature of the bias and its effect on test results has been little studied or documented. It could also be argued that the existence of whatever bias is necessary to test for some cultural understanding is appropriate in language testing, since functioning well in a foreign language involves cultural as well as linguistic knowledge.

Cargill Power (1980) also points to the fact that the test administration itself may represent cultural interference in the language testing process and thereby affect test scores. The G-TELP was administered to subjects prior to the TOEFL in this study. If subjects' first experience with a timed test occurred when they took the G-TELP, their performance would probably

have been adversely affected and their score would not represent their true level of English proficiency. The same subjects would be somewhat more experienced in taking timed tests by the time they took the TOEFL, and would therefore have performed at a level more consistent with their level of English language proficiency. The cultural interference presented by the test administration itself could have accounted for somewhat inaccurate test scores, and would have in turn affected the correlation coefficients for the two tests.

Cultural bias can also occur in the format chosen for the test questions according to Cargill-Power (1980). Since there is some variation in the types of questions included in each of the sections of the TOEFL and G-TELP, differences in test format could cause varying test performance across different native language and culture groups. This variation in test performance would, of course, influence the correlation coefficients calculated for the relationship between G-TELP and TOEFL test scores as well.

Relationship Between Subject Characteristics and Overall G-TELP and TOEFL Scores

The second hypothesis to be tested in this study was: There is no significant relationship between subjects' characteristics and overall scores on the G-TELP and TOEFL. The three subject

characteristics under consideration were age, sex, and native language.

G-TELP and TOEFL mean scores were calculated for males and females. Females earned higher G-TELP and TOEFL mean scores than males for all three levels of the G-TELP. This finding is what could be expected based on the results of performance by males and females on the TOEFL in other test administrations. For instance, in the 1987-88 TOEFL Test and Score Manual, the mean scores are given for males and females who took the test between July 1984 and June 1986. The female mean score (515) for 226,635 subjects slightly exceeded the male mean score (511) for 449,654 subjects.

As Table 2 on page 70 indicates, the highest G-TELP mean scores for males and females were earned by Level Three subjects. This is to be expected for two reasons. First, Level Three is the easiest level and therefore subjects would have the best chance of performing well on the test. Also, subjects in the Level Three group were most likely to be appropriately placed within the three G-TELP levels since it is the lowest level. Some subjects in Levels One and Two are more likely to have been improperly placed in those higher levels than subjects in Level Three. This could account for the lower mean scores for both males and females in Levels One and Two. Proper placement of

subjects in Level Three should contribute to subjects attaining higher G-TELP mean scores for that level.

The highest mean TOEFL scores were earned by G-TELP Level One males (542.11) and females (547.75). This is to be expected since Level One is the highest level of the G-TELP test series and subjects in that category would be expected to score highest on the TOEFL.

An ANOVA was calculated to determine whether a significant difference exists for males and females on overall G-TELP test scores. A significant difference was found to exist only for Level Three scores. The null hypothesis is therefore rejected for Level Three subjects and accepted for Level One and Two subjects.

An ANOVA was also calculated to determine whether a significant difference exists for males and females on overall TOEFL scores. No significant difference was found to exist for overall TOEFL scores earned by males and females for any of the three G-TELP level groups. The null hypothesis is therefore accepted for all three G-TELP level groups.

It would be interesting to compare these findings regarding differences in test performance for males and females with data gathered from other TOEFL test administrations. However, Educational Testing Service (ETS) has not published any research reports which present data regarding whether a significant difference exists between TOEFL mean scores earned by males and

females. Descriptive data has been published presenting mean scores for males and females in selected group of examinees, but no published ETS reports include ANOVA data to determine whether a significant difference exists between the two groups on overall test scores. In general, differences in test performance for males and females has not been a subject of great interest for the English language testing field.

It is interesting to note that while females in the American culture generally perform better than males on qualitative or verbal tests (which would include language tests), there was no significant difference in overall G-TELP or TOEFL scores for males and females other than for Level Three G-TELP scores. It could be hypothesized that a self-selection process occurs among males who cannot attain a certain level of language proficiency and are therefore not able to pursue higher education in the United States. It would be useful to investigate whether a significant difference exists in test scores for males and females for the immigrant population in the United States where that type of self-selection does not occur.

The second subject characteristic to be considered is age. Little if any correlation is present for subjects' age and overall G-TELP and TOEFL scores as indicated by the slightly negative correlation coefficients for all three G-TELP groups. The null hypothesis is accepted for all three G-TELP levels since little

if any relationship was found to exist between subjects' age and overall G-TELP and TOEFL scores.

It is not surprising to find that little or no relationship exists between the age variable and overall test scores because there is very little variation in subjects' age within each of the three G-TELP groups. Subjects were university bound students who were learning English because it is required for admission to United States colleges and universities, and were from a similar age group.

Again, it is not possible to compare these findings with data gained from other TOEFL administrations. ETS has published reports which present only descriptive data regarding examinees' age. No data is available regarding whether there is a significant relationship between overall TOEFL scores and subjects' age. This may be due to the fact that the relationship between subjects' age at the time of test administration and test scores has not been of major interest in the English testing field.

While it is interesting to examine the relationship between subjects' age at the time the tests were administered and test scores, it may be more relevant to an understanding of subjects' test performance to examine the age at which they first began studying English. As Carroll (1971) has noted, "It is suggested that there is a critical period for language acquisition that lasts only until about the age of puberty, with in fact a decline

of language learning ability from about the age of 5 or 6 up to puberty" (p. 109). An examination of both subjects' current age and the age at which they first began studying English would provide valuable information regarding the relationship between English proficiency test results and subjects' age.

The mean age for subjects in this study increased as the G-TELP proficiency level increased. Level One subjects show a mean age of 26.08, Level Two subjects a mean age of 23.68, and Level Three subjects a mean age of 21.98. This could be expected since the higher level students are older and could have benefited by a longer length of time to study English and increase their proficiency level.

The third subject characteristic to be considered was native language. An ANOVA was calculated to determine whether a significant difference exists for native language groups on overall G-TELP and TOEFL scores. A significant difference was found to exist only for Level One on both overall G-TELP and TOEFL scores. The null hypothesis is therefore rejected for Level One and accepted for Levels Two and Three.

There are some problems involved in interpreting the ANOVA results. The small n for several of the native language groups means that the mean score for that group is not necessarily representative of the groups' proficiency, but could be due to individual differences in test performance. It is therefore

difficult to base statements about relative proficiency by native language groups on data which may not be representative of the total native language group. Also, as was mentioned in earlier sections, placement of subjects into incorrect G-TELP levels could have influenced the mean scores earned by native language groups. Regarding the ANOVA results for differences in performance on G-TELP and TOEFL, there is no apparent explanation for why a significant difference was found to exist only for Level One test scores. However, the greater number of native languages represented and larger total n for that level may have contributed to the finding.

It would be useful to compare the mean scores attained by various native language groups for other TOEFL test administrations with the mean scores earned by subjects in this study. While there is information available regarding mean scores earned by native language groups on previous TOEFL administrations, there is a significant problem which makes such a comparison difficult.

The small n for the native language groups represented in this study means that such a comparison would not be statistically valid. For instance, for Level One TOEFL scores in Table 9 (page 76), only three of the 19 native language groups represented had more than 10 subjects, and none of the native language groups had the more than 30 n required for the results to be statistically significant. The Japanese native language group is the only one

represented in any level which was composed of 30 or more subjects for overall G-TELP and TOEFL scores. The total number of 281 subjects in this study was not large enough to permit division into three G-TELP levels and then further division within those levels into native language groups and still retain the desirable 30 or more subjects per native language group.

One study conducted by ETS researchers focuses on the performance of various native language groups on the TOEFL. Alderman (1980) investigated the performance of African, Arabic, Chinese, Japanese, Spanish, and Germanic language group subjects on the test.

He found that "the two language groups with the closest affinity to the English language, Germanic and Spanish, attained the highest TOEFL scores" (p. 12). The African group attained the next highest scores on the Structure and Reading and Vocabulary sections of the test. Alderman supposed this to be less a function of linguistic similarities with the English language than of the extensive instruction in and through English African students had received throughout primary and secondary school. In general, Alderman found that examinees from different language groups with comparable total test scores differed on their performance on specific test items, and attributed the differential item performance to linguistic similarities and dissimilarities with the English language. His study concluded,

> It seems obvious that a native language influences acquisition and performance in a second language. Less clear and perhaps unknown is the degree to which a test of proficiency in a second language accurately reflects linguistic affinity with an examinee's native language. (p. 31)

Three other studies have investigated performance on TOEFL by a few major native language groups. Dunbar (1982) analyzed data from a sample of TOEFL examinees which included eight language groups: African, Arabic, Chinese, Farsi, Germanic, Japanese, and Spanish. He found that language groups differed from each other in total TOEFL scores, with the Germanic group showing the highest mean score. Swinton and Powers (1980) found that there was a substantial variation among language groups in mean total TOEFL scores, with the Germanic group receiving the highest scores and the Farsi speakers the lowest mean scores.

In this study the Germanic language groups did not receive the highest mean TOEFL scores as they did in the previous two studies cited. However, the small n for the language groups represented in this study could have accounted for the difference in performance between studies.

There have been attempts made to evaluate examinee performance on TOEFL by groups other than native language groups. For instance, in an analysis of TOEFL examinee characteristics,

Wilson (1982) found that on the average, examinees from developed countries had higher TOEFL mean scores than examinees from developing countries. Referring to Tables 9, 11, and 13 in this paper (pages 76, 79, 81), it is clear that this was generally the case for scores in this study as well. For instance, in Table 9, eight of the ten native language groups which scored highest on the TOEFL represent developed countries. For Tables 11 and 13 there is more a mixture of native languages representing developed and underdeveloped countries throughout the ranking of performance by native language. However, the small n for many native language groups could have significantly influenced the relative ranking of performance by native language groups.

The classification systems for the diffusion of English in nations throughout the world developed by Kachru and Quirk and discussed in the literature review portion of this paper were applied to the mean score data for native language groups in Tables 9-14.

It is not surprising to note that the majority of countries represented by the native language groups present in this study's sample are expanding circle (Kachru) or EFL (Quirk) nations. The expanding circle or EFL countries are much more likely to send their students abroad to the United States for English language study and a degree than are the outer circle (Kachru) or ESL (Quirk) nations. The tradition of English language instruction

in outer circle countries like Singapore and India is well established, and English is used widely for purposes of internal communication. It is appropriate as well that no inner circle (Kachru) or ENL (Quirk) countries are represented in the lists, since English is a native language for most inhabitants of those nations.

Relationship Between Subjects' Overall and Section Scores on G-TELP and TOEFL and Scores on Post B Proficiency Test

The third hypothesis to be tested in this study was: There is no significant relationship between subjects' overall and section scores on the G-TELP and scores on the proprietary ALI English proficiency test entitled "Post B." The correlation coefficients calculated indicated there is little if any relationship between Post B and G-TELP scores for Level One subjects. For Levels Two and Three, a moderate positive correlation was found to exist between Post B and G-TELP scores. The null hypothesis is therefore accepted for Level One and rejected for Levels Two and Three. It should be noted that the Level Three results are unstable because the n is less than 30.

This is the first study which investigates the relationship between G-TELP and other language proficiency test scores, so it is not possible to compare the correlations to previous studies. However, a comparison of the test formats may help explain why

the correlations for Levels Two and Three were only moderate (.56 and .59), and why little if any correlation was found for Level One scores.

The ALI English proficiency test titled Post B consists of three parts: Listening Comprehension, Grammar, and Vocabulary and Reading. The Listening Comprehension portion includes 20 multiple choice questions. A statement is read to subjects and they are asked to choose which of three sentences is closest in meaning to the statement they just heard. Subjects are allowed 15 minutes to complete this portion of the test.

The Grammar section of the Post B test includes 30 questions. Subjects are asked to read an incomplete sentence and choose from one of the four answer choices the answer which correctly completes the sentence.

The Vocabulary and Reading section is divided into two parts. The first 30 questions are sentence completion items which require subjects to read an incomplete sentence and choose from four alternatives the word or words which best complete the sentence.

The Reading Comprehension section of the test includes 12 sentence level reading questions which ask subjects to read one or two sentences and choose one of four alternatives which best answers a question regarding the statements. The remaining eight questions are based on two short reading passages and ask subjects to choose one of four alternative statements which best answer the

questions. Subjects are allowed 50 minutes to complete the Grammar and Reading Comprehension sections of the test.

In general, the format of the Post B and G-TELP is very different. The Listening section of the Post B includes statements which are read to subjects who must then answer questions about the statements. Answering this question format requires a relatively high level skill, compared to that required to answer G-TELP questions which are based on passages in this section. The Grammar sections of the two tests are more similar in format since they both include sentence completion questions. An analysis of the grammar points being tested in this section would have to be done for each test to determine whether the section content is similar. And finally, the Reading and Vocabulary section of the G-TELP and Post B are quite different. A large portion of this Post B section is sentence completion, which includes a heavy emphasis on vocabulary; the G-TELP section includes only a few sentence completion questions.

The correlation coefficients for Post B and G-TELP scores were the lowest for Level One (-.05 and .19); no relationship was found to exist between the two test scores. There are several factors which may have influenced the Level One correlation. The first is that the most significant difference in test format and content exists between Level One and Post B. There is no Grammar section in the Level One G-TELP test and the Post B test includes

a heavy emphasis on grammar. Also, the Level One G-TELP test is the most difficult in terms of the quality and quantity of information which needs to be processed by examinees to answer the questions. This level of difficulty probably exceeds that of the Post B. And finally, the small n for all coefficients means that the results are somewhat unstable.

It is interesting to note that Post B correlates more highly with TOEFL than with G-TELP at all three levels as shown in Table 16 on page 85. This is probably due to the fact that there is more similarity in content and format between Post B and TOEFL than there is for Post B and G-TELP.

Summary

The findings regarding the three hypotheses were interpreted and discussed in this section. The relationship between G-TELP and TOEFL total and section scores was the first to be considered. In general, the moderate positive relationship discovered between total G-TELP and TOEFL scores and the Listening and Reading and Vocabulary sections of the tests were believed to be affected by the differences in test format and content and the probable improper placement of some subjects into the three G-TELP levels. The low positive relationship found for the Grammar sections of the tests was attributed to the fact that the most significant differences between test section formats and content probably exist for the Grammar sections.

The relationships discovered between subject characteristics and total G-TELP and TOEFL scores were considered as well. No significant difference in total scores was found to exist for the age and sex variables. The age range of subjects was quite restricted and contributed to this finding, and the similar scores earned by males and females were to be expected based on comparison with data from previous TOEFL administrations. And finally, there was no significant difference in performance by native language groups on the test. This finding was difficult to interpret due to the small n for all native language groups represented in the study.

The relationship between Post B and total G-TELP and TOEFL scores was considered as well. The moderate positive relationship discovered for the three test scores was attributed to differences in test formats and content.

Implications for Further Research

The results of this study indicate several areas of interest for further research. They are as follows:

1. This study represents a first attempt at comparing test results for subjects taking both the G-TELP and TOEFL. The initial findings indicate that there is a relationship between scores earned by the same subjects on both tests. The exact magnitude of that relationship is still uncertain due to the problems of accurately placing subjects in G-TELP levels and a

relatively small n for subgroups in this study. The data base begun for purposes of this study should be augmented on an ongoing basis, in order to verify the correlations calculated for this study.

2. Further research should be continued on the G-TELP itself for other types of validity. For instance, the content validity of the test should be analyzed by independent experts on an ongoing basis. It is particularly important to do this for the grammar section of the G-TELP to determine how the item level of the test relates to the item level of the TOEFL. As additional forms of the test are developed, they will need to be validated as well.

3. It would be useful to be able to predict a certain level of TOEFL performance based on a given G-TELP score, or vice versa. The initial findings of this study indicate that will be difficult to do. The common variance, or r^2, is often calculated to use as a measure of predictability. For instance, the highest correlation coefficient calculated for G-TELP and TOEFL total and section scores is .70 for Level One total scores. The r^2 for that level would be .49, which means that there is a 50% chance of predicting the TOEFL score a subject would earn based on that person's G-TELP score. This indicates a low predictive capability. Future research should be devoted to increasing the data base of G-TELP and TOEFL concurrent validity information

to aid in determining what the predictive capability is for the two tests.

4. The area of the effect of native language on English language acquisition and test performance provides excellent territory for further research. The author was initially amazed at the lack of research regarding relative performance by various native language groups on English proficiency tests. It is probable that the difficulty involved in conducting such research is that no one researcher would possess the necessary knowledge regarding several major native language groups to make the required comparisons and evaluations. In order to deal with this problem, a team of experts with expertise in major native languages of the world such as African, Arabic, Chinese, Japanese, and Spanish could be assembled. This would be a significant undertaking and should yield fascinating results.

5. The small n available for this study made it impossible to draw conclusions regarding native language groups' performance on the two tests. Further research should be done as well on the performance of various native language groups on the G-TELP and TOEFL.

6. The Post B test warrants further study as well. The test's content validity should be analyzed, and reliability studies should be updated on the test. It is important to conduct

further validity and reliability studies on the Post B prior to doing any additional comparisons with other tests.

7. Another area indicated for further research relates to the construct validity of the TOEFL and G-TELP. Native speakers of English should be given the two tests, and the results should be compared to determine what the relationship is between native speaker performance on the TOEFL and G-TELP.

8. The results of this study and others involving the G-TELP and eventually other tests of real world English should be carefully analyzed as ESL professionals prepare to meet the challenge of training second wave learners. Appropriate assessment techniques and curriculum will need to be developed for second wave learners, whose needs may be different from those of first wave learners.

In summary, suggested areas for further research include additional types of validity studies on the G-TELP, continued research regarding test performance by native language with larger n sizes, and additional types of validity studies on the Post B test.

Conclusion

This study provides valuable initial data regarding the concurrent validity of the G-TELP and TOEFL, as well as the relationship between subject characteristics and performance on the two tests. The data provided through this research, which

quantifies the relationship between TOEFL and G-TELP, will facilitate the acceptance and use of G-TELP abroad. As G-TELP is used increasingly abroad to test the real world English language proficiency of non-native speakers, it will become possible for additional groups of people to access the economic and social benefits provided by English language knowledge.

The TOEFL was designed to test the English language proficiency of an elite well educated group of non-native speakers abroad preparing to enter an American university. The author has labeled this group the first wave learners. The ESL field has focused on meeting the educational needs of this first wave group for the last three decades. The time has come to expand that focus in a new direction in order to support the development process in nations abroad. The G-TELP provides one way for the field to begin to access and focus on a new group of English language learners. These second wave learners are less well educated than the first wave learners, will not attend an American university, and have a need to demonstrate their knowledge of real world English. The G-TELP was designed to test the English language proficiency of second wave learners. The ESL field requires additional information regarding this population in order to prepare curriculum and materials which address their unique needs.

The information gained in this study will provide data which will support the recognition and eventual acceptance of the G-TELP abroad. Leaders in developing nations abroad are beginning to acknowledge the need for increased training and focus on second wave learners and are requesting assistance from ESL professionals in meeting the needs of this emerging market. The author predicts that second wave learners will be a tremendous growth market in ESL training for the next decade, and through this study and continued research in the area, hopes to be instrumental in supporting the development of second wave learners.

References

Alderman, D. L., & Holland, P. W. (1981). Item performance across native language groups on the test of English as a foreign language. Princeton, NJ: Educational Testing Service.

Alderson, J. C., Krahnke, K. J., & Stansfield, C. W., eds. (1987). Reviews of English language proficiency tests. Washington, D.C.: Teachers of English to Speakers of Other Languages.

American Psychological Association. (1985). Standards for educational and psychological testing. Washington, D.C.: Author.

Angelis, P. J., Swinton, S. S., & Cowell, W. R. (1979, October). Performance of non-native speakers of English on TOEFL and verbal aptitude tests. Princeton, NJ: Educational Testing Service.

Baldauf, R. B., & Jernudd, B. H. (1983). Language of publication as a variable in scientific communication. Australian Review of Applied Linguistics, 6, 1.

Banks, S., Ed. (1986). Political handbook of the world: 1986. New York: McGraw Hill.

Borg, W. R., & Gall, M. D. (1983). Educational research (4th ed.). New York: Longman.

Bosco, F. J. (1970). The relevance of recent psychological studies to TESOL. TESOL Quarterly, 4(1), 73.

Bowen, J. D. (1971, March). Trends in English abroad. Paper presented at the Fifth Annual TESOL Convention, New Orleans, LA.

Bowers, R. (1986). English in the world. TESOL Quarterly, 20(3), 393-410.

Briere, E. J. (1969). Current trends in second language testing. TESOL Quarterly, 3(4), 333.

Brumfit, C. (1982). English for international communication. Oxford and New York: Pergamon Press.

Buros, O. K. (1972). The seventh mental measurements yearbook. Highland Park, NJ: Gryphon Press.

Butler, F. B. (1981) Perspectives on language testing: Past, present and future. Paper presented at Nagoya Gakuin University Round Table on Languages, Linguistics and Literature, No. 4, Nagoya Gakuin University, Seta, Aichi, Japan.

Cargill-Power, C. (1980, February). Cultural bias in testing ESL. Paper presented at the Annual Meeting of the Southeastern Conference on English in the Two Year College, Birmingham, AL.

Carroll, J. B. (1971). Current issues in psycholinguistics and second language teaching. TESOL Quarterly, 5(2), 101.

Carroll, J. B. (1973). Some suggestions from a psycholinguist. TESOL Quarterly, 7(4), 355.

Cartier, F. A. (1968). Criterion-referenced testing of language skills. TESOL Quarterly, 2, 27-33.

Choi, U. (1985). English as an international language: Implications and issues. In Kim, B. W. (Ed.), Selection of speeches and papers from the International Conference on Literary and Languages. (pp. 139-148). Seoul, South Korea. Literacy and languages in Asia.

Clark, J. L. D., & Swinton, S. S. (1980, December). The test of spoken English as a measure of communicative ability in English-medium instructional settings. Princeton, NJ: Educational Testing Service.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), Educational measurement. Washington, D.C.: American Council on Education.

DeCamp, D. (1968). The current discrepancy between theoretical and applied linguistics. TESOL Quarterly, 2(1), 3.

Dizney, H. (1965). Concurrent validity of the test of English as a foreign language for a group of foreign students at an American university. Educational and Psychological Measurement, 25(4), 1129-1130.

Dizney, H. (1979). A reply from ETS. TESOL Quarterly, 13, 294-97.

Dunbar, S. B. (1982). Construct validity and internal structure of a foreign language test for several native language groups. Paper presented at the annual meeting of the American Educational Research Association, New York.

Duran, R. P., Canale, M., Penfield, J., Stansfield, C. W., & Liskin-Gasparro, J. E. (1985, February). TOEFL from a communicative viewpoint on language proficiency: A working paper. Princeton, NJ: Educational Testing Service.

Eastman, C. M. (1973). Language planning: An introduction. San Francisco: Chandler and Sharp Publishers.

Educational Testing Service (1985). TOEFL test and score manual. Princeton, NJ: Author.

Ekmekci, F. O. (1984). Sources of errors in second-language acquisition. World Language English, 3(4), 262-66.

Ervin-Tripp, S. M. (1974). Is second language learning like the first? TESOL Quarterly, 8(2), 111.

Fatham, A. K. (1976). Variables affecting the successful learning of English as a second language. TESOL Quarterly, 10(4), 433.

Fishman, J. A. (1969). National languages and languages of wider communication in the developing nations. Anthropological Linguistics, 11, 111-135.

Fishman, J. A. (1977a). Knowing, using and liking English as an additional language. TESOL Quarterly, 11(2), 157.

Fishman, J. A. (1977b). The Spread of English. Rowley, MA: Newbury House Publishers.

Gage, W., & Ohannessian, S. (1974, November). ESOL enrollments throughout the world. The Linguistic Reporter, 13-16.

Gardner, R. C. (1968). Attitudes and motivation: Their role in second-language acquisition. TESOL Quarterly, 2(3), 141.

Greenbaum, S. (Ed.) (1985a) The English language today (Richard Cureton). World Englishes, 4(2), 267.

Greenbaum, S. (1985b). 25th anniversary of the survey of English usage. World Englishes, 4(2), 261.

Hale, G. A., Angelis, P. J., & Thibodeau, L. A. (1980, December). Effects of item disclosure on TOEFL performance. Princeton, NJ: Educational Testing Service.

Hale, G. A., Stansfield, C. W., & Duran, R. P. (1983, November). Summaries of TOEFL studies, 1963-1982. Princeton, NJ: Educational Testing Service.

Hardin, G. G. (1979). English as a language of international communication: A few implications from a national point of view. English Language Teaching Journal, 4, 1-4.

Hoffman, M. S. (Ed.) (1987). The World almanac 1987. New York: Pharos Books.

Hosley, D., & Meredith, K. (1979). Inter- and intra-test correlates of the TOEFL. TESOL Quarterly, 13(2), 209-21.

Hosney, D. (1978). Performance differences of foreign students on the TOEFL. TESOL Quarterly, 12(1), 99-100.

Hudson, T. (1984). A criterion-referenced measurement approach to ESL achievement testing. Language Testing, 1(2), 171-199.

International Testing Services Center. (1985a). G-TELP test administration manual. Los Alamitos, CA: Author.

International Testing Services Center. (1985b). Questions about the G-TELP. Los Alamitos, CA: Author.

Kachru, B. B. (1984). Institutionalized second language varieties. In Sidney Greenbaum (Ed.), The English language today. Oxford: Pergamon Press.

Lee, R. R. (1972). A taxonomy of second-language behaviors. TESOL Quarterly, 6(3), 209.

Martuza, V. (1977). Applying norm-referenced and criterion-referenced measurement in education. Boston: Allyn and Bacon.

McCrum, R., Cran, W., & MacNeil, R. (1986). The story of English. New York: Viking Penguin Inc.

Millman, J. (1970). Reliability and validity of criterion-referenced test scores. In R. Traub (Ed.), New directions for testing and measurement (pp. 49-79). San Francisco: Jossey-Bass.

Nelson, C. L. (1985). My language, your culture: Whose communicative competence? World Englishes, 4(2), 243.

Oller, J. W., Bowen, J. D., Dien, T. T., & Mason, V. W. (1972, June). Cloze tests in English, Thai, and Vietnamese: Native and non-native performance. Language Learning, 22(1), 1-15.

Osanyinbi, J. A. (1975). A concurrent validity study of the West African school certificate and general certificate of education English language examination, using educational testing service's test of English as a foreign language as the criterion measure. Dissertation Abstracts International, 35, 5130A-5131A.

Popham, W. J. (1978). Criterion-referenced measurement. Englewood Cliffs, NJ: Prentice-Hall.

Quirk, R. (1969). English today: A world view. TESOL Quarterly, 3(1), 23.

Quirk, R., & Widdowson, H. G. (Ed.). (1985). English in the world: Teaching and learning the language and literatures. Cambridge: Cambridge University Press.

Rice, F. A. (1962). Study of the role of second languages in Asia, Africa, and Latin America. Arlington, VA: Center for Applied Linguistics.

Ridler, N. B., & Pons-Ridler, S. (1984). Universal languages: An economic rationale. World Language English, 4(1), 1-4.

Shuy, R. W. (1969). The relevance of socio-linguistics for language teaching. TESOL Quarterly, 3(1), 13.

Starr, S. F. (1978, May). English dethroned. Change, 10(5), 26-31.

Swales, J. (1985). English as the international language of research. Regional Language Center Journal, 16(1), 1-7.

Swinton, S. S., & Powers, D. E. (1980). Factor analysis of the Test of English as a Foreign Language for several language groups. Princeton, N.J.: Educational Testing Service.

Trifonovitch, G. J. (1978). English as an international language: An attitudinal approach. (ERIC Document).

United Nations. (1984). Statistical yearbook, 1983/84. New York: United Nations.

United Nations. (1985). Statistical yearbook, 1985. New York: United Nations.

Wilson, K. M. (1982). A comparative analysis of TOEFL examinee characteristics, 1977-79. Princeton, NJ: Educational Testing Service.

Woodward, P. E. (1980, November). The test of English for international communication (TOEIC). Paper presented at the meeting of the English-speaking Union Conference, London, England.

Za'rour, G. I., & Buckingham, T. (1969). Some factors affecting improvement in proficiency in English as a second language. TESOL Quarterly, 3(1), 37.