




Theses and Dissertations

1988

The Test of English as a Foreign Language Sample Test as a Measure of Adolescent Language Ability

Paul Gardiner Osborn
Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Educational Methods Commons](#), [Linguistics Commons](#), and the [Mormon Studies Commons](#)

BYU ScholarsArchive Citation

Osborn, Paul Gardiner, "The Test of English as a Foreign Language Sample Test as a Measure of Adolescent Language Ability" (1988). *Theses and Dissertations*. 5003.
<https://scholarsarchive.byu.edu/etd/5003>

This Thesis is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

LB
1103.02
.082

The Test of English as a Foreign Language Sample Test
as a Measure of Adolescent Language Ability

A Thesis

Presented to the

Department of Educational Psychology

Brigham Young University

In Partial Fulfillment

of the Requirements for the Degree

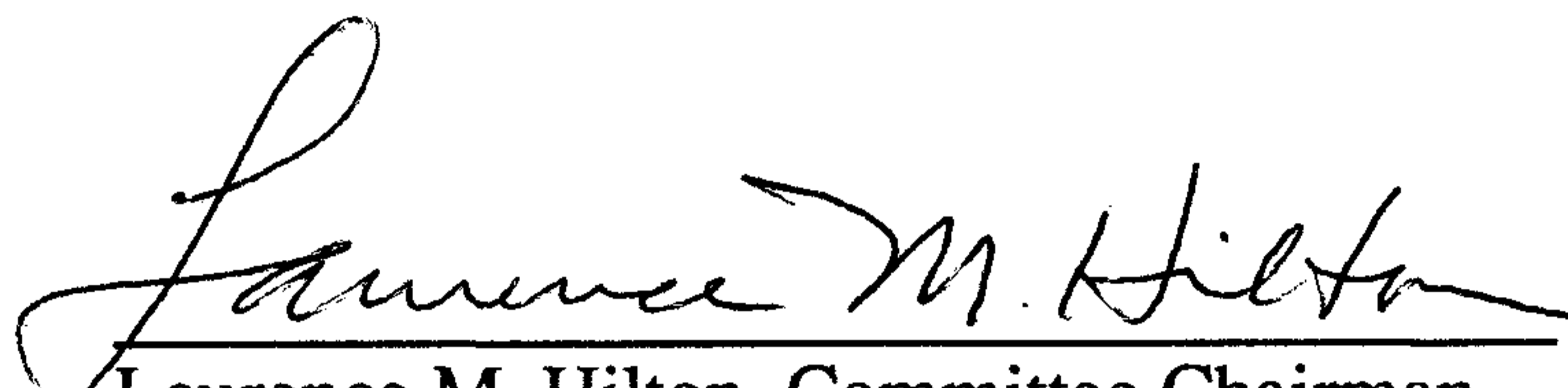
Master of Science

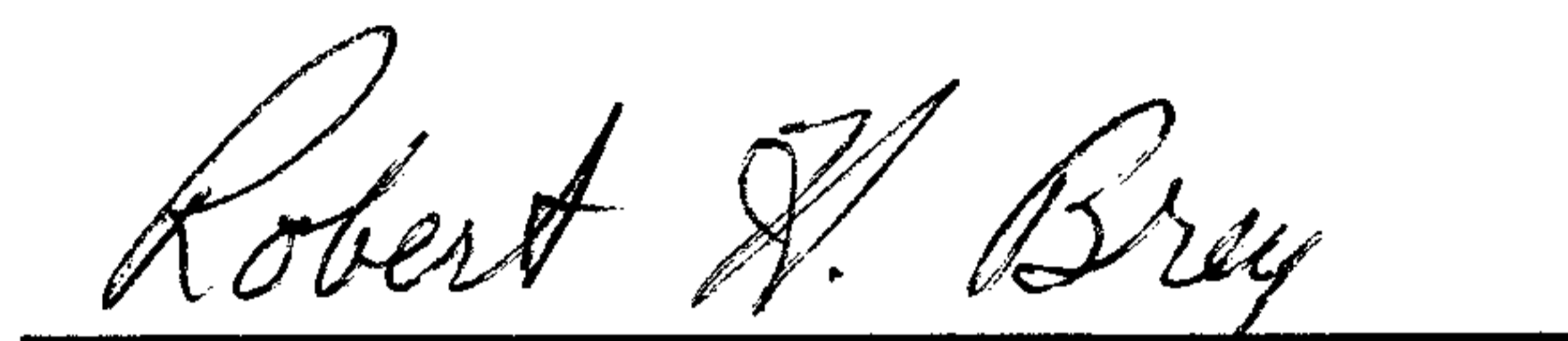
by

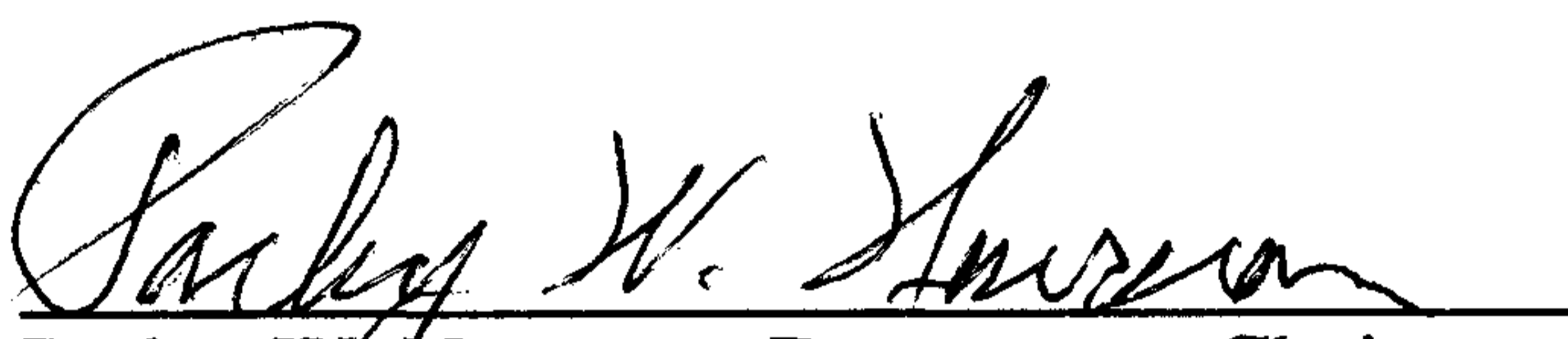
Paul Gardiner Osborn

August, 1988

This thesis, by Paul Gardiner Osborn, is accepted in its present form by the Department of Educational Psychology of Brigham Young University as satisfying the thesis requirement for the degree of Master of Science.


Laurence M. Hilton, Committee Chairman


Robert H. Brey, Committee Member


Parley W. Newman, Department Chairman

7-11-88
Date

ACKNOWLEDGMENTS

I would like to express my appreciation to Dr. Laurence M. Hilton for his invaluable assistance in every stage of this investigation and to Wendy Cloward and Michelle Coles for their assistance in the collection of the data.

Gratitude is also expressed to the Church Educational System of the Church of Jesus Christ of Latter-Day Saints and to the Timp View High School Seminary for their help in obtaining subjects for this study.

I would also like to express sincere gratitude to my dear wife, LaDawn S. Osborn, whose love and support made the completion of this project possible.

Table of Contents

	Page
Acceptance Page	ii
Acknowledgments.....	iii
List of Appendices	v
List of Tables.	vi
Journal Article	1
Title Page	1
Abstract.....	2
Method.....	9
Subjects	9
Instruments.....	9
Procedure.....	11
Data Analysis	12
Results.....	12
Discussion.....	23
References	27
Appendix A.....	32
Appendix B.....	50
Appendix C.....	51
Abstract.....	54

List of Appendices

Appendix	Page
A. Review of the Literature	32
References	45
B. Subject Background Questions.....	50
C. Informed Consent Document.....	51

Table	Page
1. Means and standard deviations for the TOEFL-ST and the PPVT-R.....	14
2. Correlations of test scores with sex and age	16
3. Means and standard deviations for the FLTA.....	17
4. Correlations among language test total scores	19
5. Total and subtest correlations among language tests.....	21

The Test of English as a Foreign Language Sample Test
as a Measure of Adolescent Language Ability

Paul G. Osborn

Department of Educational Psychology

Brigham Young University

Running Head: MEASURES OF ADOLESCENT LANGUAGE

Abstract

Relative performance on the Test of English as a Foreign Language Sample Test (TOEFL-ST) was explored in sixty native English speaking high school students. Subjects also were administered the Fullerton Language Test for Adolescents and the Peabody Picture Vocabulary Test - Revised. The TOEFL-ST was not difficult for this population, indicating that TOEFL tests taken by foreign speaking college students probably assess a level of native English competency well below the high school level. The three tests, including subtests, appear to measure a wide array of subdomains of language competency. The data do not support the conclusion that any of these tests could be substituted for the others in assessing language competency.

The Test of English as a Foreign Language Sample Test
as a Measure of Adolescent Language Ability.

New and more efficient ways of assessing language development and competency continually are being sought. One instrument that has been used to assess English language proficiency is the Test of English as a Foreign Language (TOEFL). Colleges and universities throughout the United States and Canada use the TOEFL as a screening instrument to measure English language proficiency of foreign students seeking admission (Ayers & Peters, 1977; Bostic, 1981; Duran, Canale, Penfield, Stansfield, & Liskin-Gasparro, 1985; ETS, 1987). A TOEFL Committee of Examiners Policy Council, a panel of specialists in linguistics, language testing and teachers of English as a foreign or second language continually reviews the TOEFL to ensure that the most current trends and methodologies in the field are reflected in the test's content (ETS, 1987). A new equivalent form of the TOEFL is developed for each consecutive administration.

The TOEFL Program publishes official preparation and study materials. These materials are developed to aid students in understanding the format of the test and the special linguistic skills measured by the TOEFL (ETS, 1987). The TOEFL Sample Test (ETS, 1985) contains questions used at previous TOEFL administrations, and is representational of the TOEFL. The TOEFL

Sample Test (TOEFL-ST) is approximately one-half the length of the actual TOEFL.

The Educational Testing Service (ETS, 1987) estimates the developmental difficulty of the TOEFL to be at the 11th grade level. The test evaluates receptive language ability in the areas of grammar, vocabulary, and usage (Duran et al., 1985). The TOEFL currently is administered in three sections which are timed and presented separately. In multiple choice format, the three sections assess receptive abilities in listening comprehension, structure and written expression, and vocabulary and reading comprehension (Alderman, 1982; Clark, 1977; ETS, 1987).

Because the TOEFL is so widely used as a method of screening foreign students, it has been the subject of much research directed toward groups for whom English is a second language. (Abadzi, 1976; Andalib, 1976; Angelis, Swinton, & Cowell, 1979; Ayers & Peters, 1977; Carlson, Bridgeman, Camp, & Waanders, 1985; Dizney, 1965; Gharavi, 1977; Hosley, 1978; Komvichayungyuen, 1977; Odunze, 1982; Wilcox, 1975). Few studies have looked at the usefulness of the TOEFL with adolescent and young adult native speakers of English (Angoff & Sharon, 1971; Clark, 1977; Johnson, 1977).

The TOEFL is presented by its authors to be a robust measure of English proficiency, measuring the non-native prospective college student's ability in several areas of English language proficiency (ETS,

1987). The ETS contends that high scores on the TOEFL reflect a competency level in English that signifies readiness to cope with English language demands needed to pursue studies at United States colleges and universities. The TOEFL is not intended for use as a test of native English language proficiency. The ETS states that "the TOEFL test is recommended for students at the eleventh-grade level or above; the test content is considered inappropriate for younger students" (ETS, 1987, p.6). This is interpreted by many test users to represent a statement that the TOEFL reflects at least 11th grade or high school level equivalency of English proficiency within the areas tested. In fact, the actual levels of developmental native English competency or academic achievement assessed by the TOEFL have never been determined.

In one study (Clark, 1977), the TOEFL was administered to 88 native English speaking high school seniors prior to graduation. Assuming the TOEFL was developed to present language tasks similar to those encountered by foreigners in the United States, Clark postulated that an educated English speaker would be capable of meeting or exceeding the test's standard of language performance. If the native speakers were found to have no difficulty with the test material, the assumption was that the TOEFL test questions were appropriate for determining English competency in non-native speakers. The results of the study indicated that the TOEFL, on the

whole, was easy for this population. Total test scores indicated that the native English speaking group performed extremely well on the test, and the Clark data are the apparent basis for the ETS's statement that the TOEFL is recommended for students at the 11th grade level or above. Clark also conjectured that due to ceiling effects seen in these native English speaking graduating high school seniors, any use of the TOEFL to measure language proficiency among native speakers would be virtually useless (Clark, 1977).

Few TOEFL studies have been conducted on handicapped populations. Ragosta and Nelson (1986) used the TOEFL to measure the language proficiency of profoundly hearing impaired native English language college students. The TOEFL was presented to 26 college students, 18 of whom were deaf from birth and 24 of 26 were deaf by at least age 15. All were proficient in American Sign Language. Test instructions were spoken, and listening comprehension test items were played on a tape recorder. An interpreter simultaneously signed the same words. The students generally scored in the 2nd quartile, better than chance performance but considerably lower than foreign college students' scores. Ragosta and Nelson concluded that the TOEFL appeared to be a valid measure of relative English language proficiency in deaf college students. This study established the usefulness of the TOEFL in evaluating language

proficiency in communication disordered native English speaking populations.

Experienced examiners have reported that assessment of adolescent language is a difficult task. This is in part due to noncompliant test taking behaviors often seen in this age group (Emerick & Haynes, 1986), and in part to inadequacies in existing published assessment instruments (Compton, 1984; Duran et al., 1985). Several language tests currently used to test adolescent language abilities include the Test of Adolescent Language (TOAL) (Hammill, Brown, Larson, & Wiederholt, 1980), the Fullerton Language Test for Adolescents, 2nd Edition (FLTA) (Thorun, 1986), the Peabody Picture Vocabulary Test - Revised (PPVT-R) (Dunn, 1981), the Clinical Evaluation of Language Functions (CELF) (Semel & Wiig, 1980), and the Screening Test of Adolescent Language (STAL) (Prather, Breecher, Stafford & Wallace, 1980).

Lieberman, Heffron, West, Hutchinson, and Swen (1987) reported that although the content and procedures of administering the TOAL, FLTA, CELF, and STAL differ from one another, student performance correlations between them were moderate to high. The TOLD assesses the adolescent's receptive and expressive vocabulary and grammar in written and spoken forms. Eight subtests are included for the purpose of measuring interactions among these parameters. The FLTA consists of eight norm referenced subtests that are designed

to differentiate language impaired adolescents from normal language adolescents, and to identify strengths and weaknesses in language comprehension and production of these individuals. Two subtests assess receptive language and six assess expressive language. The subtests are not sequence dependent (Compton, 1984).

The PPVT-R is a standardized receptive vocabulary assessment tool that assesses receptive knowledge of nouns, verbs, and adjectives. The test is easy to administer and score. Scoring procedures provide the examiner with an age score equivalent based on norms from age 2 years to 40 years (Dunn, 1981).

Although past TOEFL research has examined concurrent validity of the test with other measures of English language proficiency for foreign speakers (Carlson, 1985; Pack, 1971; Perkins & Pharisk, 1980), it has not been compared to native language assessment instruments. The purposes of this study are: first, using the TOEFL Sample Test (a representative short form of previous TOEFL tests), to examine the validity of the commonly held assumption that the TOEFL reflects an 11th grade or higher level of English language proficiency in the domains tested; second, to discover the degree to which the TOEFL Sample Test (TOEFL-ST) can differentiate students at grades 10, 11, and 12; and third, to explore the extent to which TOEFL-ST scores correlate with performance on the FLTA and the PPVT-R in a

The TOEFL-ST as a Measure of Adolescent
Language Ability

Paul Gardiner Osborn

Department of Educational Psychology

M. S. Degree, August 1988

ABSTRACT

Relative performance on the Test of English as a Foreign Language Sample Test (TOEFL-ST) was explored in sixty native English speaking high school students. Subjects also were administered the Fullerton Language Test for Adolescents, and the Peabody Picture Vocabulary Test - Revised. The TOEFL-ST was not difficult for this population, indicating that TOEFL tests taken by foreign speaking college students probably assess a level of native English competency well below the high school level. The three tests, including subtests, appear to measure a wide array of subdomains of language competency. The data do not support the conclusion that any of these tests could be substituted for the others in assessing language competency.

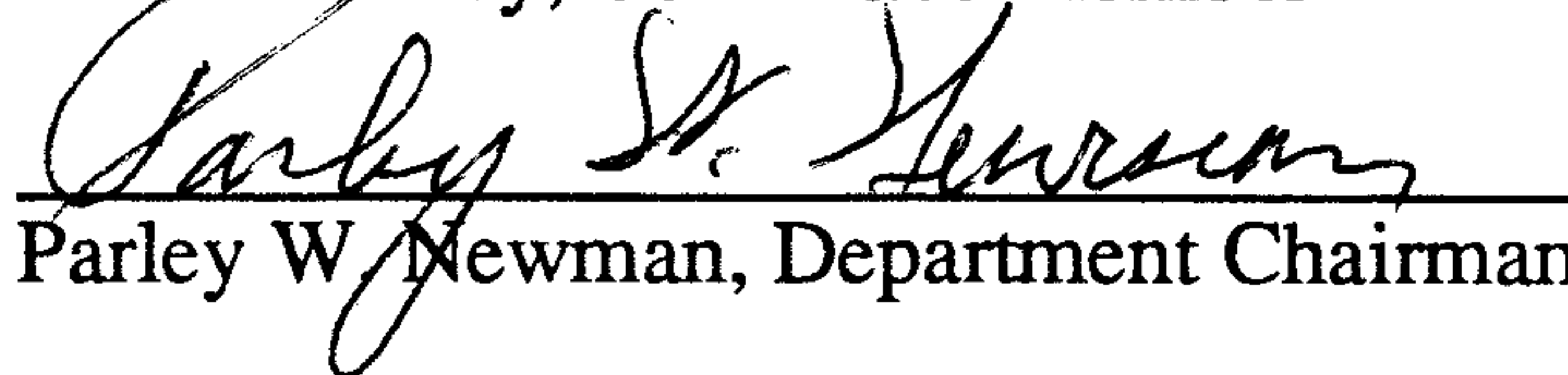
COMMITTEE APPROVAL:



Laurence M. Hilton, Committee Chairman



Robert H. Brey, Committee Member



Parley W. Newman, Department Chairman

sample of 60 native General American English speaking high school students.

Methods

Subjects

Subjects were sixty native English speaking public high school students with no history of hearing, language or learning problems. In addition to being enrolled in high school classes, subjects also attended a released-time religious education seminary class which was held during regular school hours in a building adjacent to their high schools. Subjects were volunteers from these seminary classes.

Each subject passed an air conduction hearing screening test prior to inclusion in the study consisting of tones presented to each ear at 500, 1000, 2000, and 4000 Hz at 15 dB HL (ANSI, 1970). Using a background history questionnaire (see appendix), individuals with a positive history of previous language disorder, hearing loss, learning disorders, or other than academic exposure to non-native language influences were excluded from the study. All subjects were required to submit a signed informed consent document prior to participation (see appendix). Three groups of twenty adolescents, ten males and ten females, representing grades 10, 11, and 12 participated.

Instruments

The TOEFL-ST consists of three sections including listening comprehension, structure and written expression, and vocabulary and

reading comprehension. Each section is timed and requires approximately 20 minutes to complete. For all sections, times listed include time spent reading the directions. In the listening comprehension section, the student is presented previously recorded sentences or short passages. The student is then required to answer questions about what they had just heard. For example, the student hears, "Mary swam out to the island with her friends." The student then reads (A) Mary outswam the others. (B) Mary ought to swim with them. (C) Mary and her friends swam to the island. (D) Mary's friend owned the island, and selects the item which is closest in meaning to the sentence they heard. This section allows 20 minutes for completion. In the 20 item structure and written expression section, the student is required to complete incomplete sentences with an appropriate given word (cloze procedure) or to identify an inappropriately used word. For example, the student would read "Vegetables are an excellent source _____ vitamins." The student would then select from (A) of, (B) has, (C) were and (D) that. This section allowed 15 minutes for completion. In the 30 item vocabulary and reading comprehension section of the TOEFL-ST, a short reading passage is presented. After reading the selection the student is required to answer multiple choice questions. This section allows 25 minutes for completion. A raw score was calculated for each section by subtracting the number of items missed from the total number of items possible for each section.

The FLTA consists of eight subtests titled Auditory Synthesis, Morphology Competency, Oral Commands, Convergent Production, Divergent Production, Syllabication, Grammatic Competency, and Idioms. Subtests are not timed but each requires approximately five minutes to administer. Following FLTA standardization protocol, all subtests were individually administered. Raw scores and percent correct scores were obtained for each subtest. The FLTA scoring manual (Thorun, 1986), contains tables for comparing raw scores to means and standard deviations. These tables were consulted for descriptive and interpretive purposes.

The PPVT-R consists of 175 stimulus items and requires approximately 15 minutes to administer. A starting point is recommended based on the chronological age of the subject. Basal and ceiling items were determined for each subject. Basal is defined as the highest 8 consecutive correct responses. The ceiling is the lowest level at which 6 of 8 consecutive responses are missed. A raw score is obtained by subtracting the number of errors obtained between the basal and the ceiling from the ceiling item. The raw score also can be converted into a standard score equivalent, a stanine score, a percentile rank or an age equivalent.

Procedure

Each subject was administered the PPVT-R, the TOEFL-ST and the FLTA by a single examiner. Each test was presented using the

standard procedures outlined in the test manuals. The PPVT-R was administered first to all prospective subjects and was scored immediately. Subjects scoring more than one standard deviation below the mean for their chronological age were screened from the study and were not administered the remaining tests. The presentation sequence of the FLTA and the TOEFL-ST was counterbalanced across subjects. The entire procedure was completed in approximately two hours.

Data Analysis

Raw scores were obtained for each subtest of the TOEFL Sample Test, the FLTA, and the PPVT-R. Descriptive statistics were computed including percentages, means, standard deviations, as well as standard score equivalents on the standardized measures. Pearson's Product Moment Correlations among subtest scores, total test scores and age were computed.

Results

Table 1 reports mean scores and standard deviations obtained for grades 10, 11, and 12 for each of the TOEFL-ST subtests, TOEFL total scores, and the PPVT-R.

No norms have been published for the TOEFL Sample Test used in this study. There were 75 points possible on the TOEFL-ST. When total TOEFL-ST scores in Table 1 are converted to percentages, 10th, 11th, and 12th grade subjects scored 91 percent, 96 percent, and 96

percent correct respectively. Subtest 1 scores converted to percentages showed 10th, 11th, and 12th grade percent correct performances to be 94 , 97 , and 97 percent respectively. Converted subtest 2 scores showed 10th, 11th, and 12th grade percent correct

Insert Table 1 About Here

performances to be 92, 98, and 97 percent correct respectively. Tenth graders scored 89 percent correct, 11th graders 94 percent, and 12th graders 92 percent on Subtest 3.

On the PPVT-R, of the sixty subjects tested, 39, or 65 percent, scored within plus or minus one standard deviation of the mean for their age. Ten subjects, 17 percent, scored within one standard deviation below the mean for their chronological age, while 29 subjects, 48 percent, scored within one standard deviation above the mean. Eighteen subjects, 30 percent of the sample, scored two standard deviations above the mean, and the remaining three subjects, 5 percent, scored three standard deviations above the normative mean for their ages. Overall, 17 percent of the subjects scored below the age appropriate mean on the PPVT-R, 83 percent above.

Table 2 reports Pearson's Product Moment Correlation Coefficients for age of the subject with total and subtest scores obtained on the TOEFL Sample Test, the PPVT-R, and the FLTA. In

Table 1

Means and Standard Deviations for the TOEFL-ST and the PPVT-R

	10th Grade		11th Grade	
	Mean	SD	Mean	SD
TOEFL-ST				
Total	68.30	4.55	71.80	3.39
Subtest 1	23.40	1.31	24.20	0.88
Subtest 2	18.40	1.46	19.50	0.83
Subtest 3	26.60	2.65	28.20	2.70
PPVT-R				
Total	149.00	9.70	156.00	7.36
	12th Grade		Combined	
	Mean	SD	Mean	SD
TOEFL-ST				
Total	71.70	4.27	70.60	4.35
Subtest 1	24.20	1.06	23.90	1.15
Subtest 2	19.40	0.88	19.10	1.19
Subtest 3	27.70	3.98	27.50	3.19
PPVT-R				
Total	154.00	10.60	153.00	9.70

addition, point biserial correlations for sex of the subject with total and subtest scores obtained on the TOEFL Sample Test, the PPVT-R, and the FLTA are also reported. No significant correlations were found between any of the obtained scores, sex, and age. Subjects included 30 males and 30 females. Subjects ranged in age from 15-6 to 19-0 years. Two subtests of the FLTA were found to correlate

Insert 2 Table About Here

significantly with sex of the subject: subtest 1 ($r = .40$, $p < .01$), auditory synthesis of phonemes and syllables, and subtest 2 ($r = .34$, $p < .05$), morphology competency ("Use this word in a sentence.").

Table 3 reports mean scores and standard deviations obtained on the FLTA by subjects in grades 10, 11, and 12. For the purpose of comparison, raw scores for each subject were collapsed across subtests to obtain a total FLTA score. When the mean of this total score is

Insert Table 3 About Here

compared with the grand mean of the subtest means on the Thorun (1986) FLTA normative performance profile, it can be seen that subjects in the present study scored higher as a group than the standardization sample (these means were 203.60 and 192.00 respectively). A standard deviation for total score in the

Table 2

Correlations of Test Scores With Sex and Age

	Sex	Age
Age	-.173	
TOEFL-ST		
Total	.073	.224
Subtest 1	.000	.198
Subtest 2	.071	.217
Subtest 3	.026	.108
PPVT-R	-.152	.117
FLTA		
Total	.112	.168
Subtest 1	.400	.027
Subtest 2	.069	-.063
Subtest 3	-.119	.003
Subtest 4	-.003	.125
Subtest 5	-.032	.177
Subtest 6	.070	-.004
Subtest 7	.056	.043
Subtest 8	.344	.054

* $p < .05$ ** $p < .01$

Table 3

Means and Standard Deviations for the FLTA

	<u>Grade</u>					
	10th		11th		12th	
	Mean	SD	Mean	SD	Mean	SD
Total	198.95	13.13	205.50	15.79	206.45	16.50
Subtest 1	17.90	2.31	18.20	1.63	18.70	0.92
Subtest 2	18.00	1.59	18.00	1.57	18.30	1.25
Subtest 3	19.40	1.05	19.20	0.81	18.70	0.59
Subtest 4	40.00	5.17	44.00	4.71	43.10	4.84
Subtest 5	56.00	5.36	56.20	13.10	56.50	10.40
Subtest 6	18.00	2.20	19.00	1.00	18.20	2.82
Subtest 7	18.00	1.56	18.20	1.01	18.60	0.88
Subtest 8	11.70	3.85	12.80	2.98	13.50	2.96
	<u>Combined</u>					
	Mean	SD		Mean	SD	
Total	203.60	15.32	Subtest 5	56.20	9.94	
Subtest 1	18.30	1.72	Subtest 6	18.40	2.15	
Subtest 2	18.10	1.46	Subtest 7	18.30	1.19	
Subtest 3	19.40	0.58	Subtest 8	12.70	3.32	
Subtest 4	42.40	5.13				

standardization sample was not published by the test's author. When individual FLTA subtest means obtained in the current study were compared with means and standard deviations in the Thorum standardization performance profile, all but one were found to be within one standard deviation above the mean. The mean on FLTA subtest five, a five item divergent production verbal fluency task, was within one standard deviation below the normative mean. It should be remembered that students with any history of hearing, language or learning problems were excluded from the study.

Table 4 reports Pearson's Product Moment Correlation Coefficients for total scores among the TOEFL-ST, the FLTA, and the PPVT-R. Coefficients among these total scores on the three tests

Insert Table 4 About Here

revealed moderate but significant correlations. The correlation coefficient between the TOEFL-ST and the FLTA total scores was .478. Correlations among the TOEFL-ST total scores, FLTA total scores and the PPVT-R demonstrated moderately positive correlations ranging from .536 to .602. All of these correlations were significant at the .01 level.

Table 5 reports Pearson's Product Moment Correlation Coefficients among subtests of the TOEFL-ST and FLTA and total

Table 4

Correlations Among Language Test Total Scores

	PPVT-R	TOEFL-ST
TOEFL-ST	.602**	
FLTA	.536**	.478**

** $p < .01$

scores on the TOEFL-ST, FLTA, and PPVT-R. Of the correlations computed among test scores, 32 were significant at the .01 level, 10 were significant at the .05 level, and 41 did not reach significance. TOEFL-ST total scores correlated most highly with PPVT-R scores. Several of the subtests of the TOEFL-ST and the FLTA were found to be moderately correlated. TOEFL subtest 1, listening comprehension, and FLTA subtest 8; defining idioms, exhibited the the strongest between subtest correlation, $r = .526$. TOEFL subtest 2, structure and written expression, correlated most highly, $r = .447$, with FLTA subtest 7, grammatic competency. TOEFL subtest 3, vocabulary and reading comprehension, showed the greatest strength of correlation,

Insert Table 5 About Here

$r = .395$, with FLTA subtest 4, convergent production. Other statistically significant correlations among TOEFL and FLTA subtests ranged from $r = .357$ to $r = .479$.

Scores on the Peabody Picture Vocabulary Test - Revised correlated significantly ($p < .01$) with TOEFL subtests 1, 2, and 3 and FLTA subtests 2, 4, 6, and 8. Of these, TOEFL subtest 1 ($r = .604$) and FLTA subtest 4 ($r = .577$) were most highly correlated with the PPVT-R.

Table 5

Total and Subtest Correlations Among Language Tests

	PPVT-R	FLTA Subtests				
		1	2	3	4	5
FLTA						
Total	.536**					
Subtest 1	.122					
Subtest 2	.427**	.323*				
Subtest 3	.277*	.162	.047			
Subtest 4	.577**	.097	.386**	.096		
Subtest 5	.164	.007	-.005	.030	.217	
Subtest 6	.364**	.015	.219	-.188	.303*	.171
Subtest 7	.319	.347*	.428**	.161	.286*	-.192
Subtest 8	.420**	.278*	.476**	.271	.466**	.081
TOEFL-ST						
Total	.602**	.223	.290*	.266	.510**	.106
Subtest 1	.604**	.030	.146	.234	.479**	.045
Subtest 2	.493**	.193	.243	.233	.368**	.140
Subtest 3	.426**	.167	.256	.152	.395**	.132

(table continues)

	FLTA Subtests				TOEFL-ST Subtests	
	6	7	8	Total	1	2
FLTA						
Subtest 7	.052					
Subtest 8	.232	.284*				
TOEFL-ST						
Total	.158	.468**	.517**	.478**		
Subtest 1	.347	.256	.526**	.402**		
Subtest 2	.026	.447**	.422**	.402**	.466**	
Subtest 3	.140	.388**	.357**	.397**	.458**	.420**

* $p < .05$ ** $p < .01$

Discussion

Using the TOEFL-ST (a representative short form of previous TOEFL tests), the primary purpose of this study was to examine the validity of the commonly held assumption that the Test of English as a Foreign Language reflects an 11th grade level or higher of English language proficiency in the domains tested. A second purpose was to discover the degree to which the TOEFL Sample Test can differentiate students at grades 10, 11, and 12. A third purpose was to explore the extent to which TOEFL-ST scores correlate with performance on the FLTA and the PPVT-R.

Clark's (1977) conclusion that native English speakers with an 11th grade level of English competency would meet or exceed the TOEFL's standard of language performance is consistent with the findings of this study. However, to conclude that excellent performance on TOEFL tests represents an underlying English language competency typical of 11th grade native English speakers probably is not justified. In the present study, no clear differentiation by sex, age or grade was seen in TOEFL Sample Test scores for subjects representing grades 10, 11 and 12. Students at all grade levels averaged above 90 percent correct on the TOEFL-ST. Although these data were collected during the seventh and eighth months of the school year, results indicated that 10th grade subjects had little difficulty with the TOEFL-ST. Given the fact that the 10th graders' overall mean on the TOEFL-ST was 91

percent correct, TOEFL tests probably assess a level of native English language competency well below the high school level. Certainly, college administrators considering foreign applicants for admission to United States colleges and universities on the basis of TOEFL scores should be aware that high scores on a TOEFL administration may reflect only a very limited estimate of English language competency, and probably reflect at best a junior high school level of English language competency in the foreign student.

No significant correlations with sex or age were found in the data. Many moderate to low but statistically significant correlations (none over .60) were found among the performance measures, i.e. among total scores and sub-test scores. No clear patterns of association were evident among the various scores. Beyond concluding the obvious fact that all are measures of the English language proficiency of the students tested, little else can be said. The three tests, including subtests, appear to measure a wide array of subdomains of language competency. The data do not support the conclusion that any one of these tests (or their individual subtests) could be substituted for the others in assessing the language competencies of the subjects included in this study.

Sample means generally higher than the age level standardization means indicate that the sample studied was a more homogeneous group than the norming samples employed in the standardization of the

normed tests (Thorun, 1986). The population would best be described as being average to above average on the test instruments used.

It is possible that sampling restrictions excluded a balanced representation of students with relatively lower ability and were responsible for the observed trends in the data to be shifted toward higher scores on both the Fullerton Language Test for Adolescents and the Peabody Picture Vocabulary Test - Revised. The high school students sampled in the present study were not a randomly selected representative sample. Perhaps most importantly, students with any history of hearing disorders, language disorders or learning problems were excluded from the study in an effort to sample native English speakers with normal language competency. Secondly, subjects volunteered to participate, and the majority ($n = 42$) were contacted during regular school hours while attending a released-time religious education seminary class which convened in a building adjacent to their high school. The remaining eighteen participants also were seminary students, but were contacted directly. It is recommended that interpretations of data obtained in this study and their generalization to other populations take these sampling restrictions into account.

As a research tool, the TOEFL already has been shown to be useful in assessing the relative language competency of native English

speaking deaf college students (Ragosta and Nelson, 1986). The possibility still exists that the TOEFL or the TOEFL-ST may prove useful in differentiating other populations of native English speakers with communication disorders. Further research is needed to explore the potential usefulness of the TOEFL-ST in assessing language competency in native English speaking populations and to establish age level norms and grade level equivalency norms within normal and communicatively disordered populations before the TOEFL Sample Test could be used as a reliable diagnostic tool.

References

- Abadzi, J. (1976). Evaluation of foreign students' admission procedures used at the University of Alabama (Doctoral dissertation, University of Alabama, 1975). Dissertation Abstracts International, 36, 7754A.
- Alderman, D. L. (1982). Language proficiency as a moderator variable in testing academic aptitude. Journal of Educational Psychology, 74, 580-587.
- American National Standards Institute (1970). Specifications for audiometers (ANSI S3.6-1969). New York: ANSI.
- Andalib, A. A. (1976). The academic success of undergraduate Iranian students in selected Texas Universities (Doctoral dissertation, East Texas State University, 1975). Dissertation Abstracts International, 36, 4881A.
- Angelis, P., Swinton, S., & Cowell, W. (1979). The performance of non-native speakers of English on TOEFL and verbal aptitude tests. (Report No. ETS 3) Princeton, NJ: Educational Testing Service .
- Angoff, W., & Sharon, A. (1971). A comparison of scores earned on the TOEFL by native American college students and foreign applicants to U.S. colleges. TESOL Quarterly, 5, 129-136.
- Ayers, J. B., & Peters, R. M. (1977). Predictive validity of the Test of English as a Foreign Language for Asian graduate students in

- engineering, chemistry, or mathematics. Educational and Psychological Measurement, 37, 461-463.
- Bostic, M. L. (1981). A correlational study of academic achievement and the Test of English as a Second (sic) Language (TOEFL) (Doctoral dissertation, University of Chicago, 1981).
Dissertation Abstracts International, 42, 468A.
- Carlson, S. B., Bridgeman, B., Camp, R. & Waanders, J. (1985).
Relationship of admissions test scores to writing performance of native and nonnative speakers of English. (Report No. ETS 19) Princeton, NJ: Educational Testing Service.
- Clark, J. L. D. (1977). The performance of native English speakers on the Test of English as a Foreign Language. (Report No. ETS 1) Princeton, NJ: Educational Testing Service.
- Compton, C. (1984). A guide to 75 tests for Special Education. Belmont, California: Pitman Learning.
- Dizney, H. (1965). Concurrent validity of the Test of English as Foreign Language for a group of foreign students at an American University. Educational and Psychological Measurement, 25, 1129-1131.
- Dunn, L. M. (1981). The Peabody Picture Vocabulary Test - Revised. Circle Pines, MN: American Guidance Service.
- Duran, R. P., Canale, M., Penfield, J., Stansfield, C. & Liskin-Gasparro, J. (1985). TOEFL from a communicative viewpoint

- on language proficiency: A working paper. (Report No. ETS 17). Princeton, NJ: Educational Testing Service.
- Educational Testing Service (ETS), (1987). TOEFL Test and Score Manual. Princeton, NJ: Author.
- Educational Testing Service (ETS) (1985). Test of English as a Foreign Language (TOEFL) Sample Test. Princeton, NJ: Author.
- Emerick, L. & Haynes, W. (1973). Diagnosis and evaluation in speech pathology (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Gharavi, E. (1977). Admission of foreign graduate students: An analysis of judgments by selected faculty and administrators at North Texas State University (Doctoral dissertation, North Texas State University, 1977). Dissertation Abstracts International, 38, 1148A.
- Hammill, D. D., Brown, V. L., Larson S. C., & Wiederholt, J. L. (1980). Test of Adolescent Language. Austin, TX: Pro-Ed.
- Hosley, D. (1978). Performance difference of foreign students on the TOEFL. TESOL Quarterly, 12, 99-100.
- Johnson, D. (1977). The TOEFL and domestic students: Conclusively inappropriate. TESOL Quarterly, 11, 79-86.
- Komvichayungyuen, N. (1977). Test of English as a Foreign Language as a predictor of actual language proficiency.

(Doctoral dissertation, Florida State University, 1977).

Dissertation Abstracts International, 38, 6666A.

Lieberman, R. J., Heffron, A., West, S., Hutchinson, E., & Swen, T.

(1987). A comparison of four adolescent language tests.

Language, Speech, and Hearing in Schools, 18, 250-265.

Odunze, O. J. (1982). Test of English as Foreign Language and first year G.P.A. of Nigerian students. (Doctoral dissertation,

University of Houston, 1981). Dissertation Abstracts

International, 42, 3419A-3420A.

Pack, A. (1972). A comparison between TOEFL and Michigan Test

scores and student success in (1) Freshman English and (2)

Completing a college program. TESOL Reporter, 9, 1-7.

Perkins, K., & Pharisk, K. (1980). TOEFL scores in relation to

standardized reading tests. Research in Language Testing,

142-146.

Prather, E., Breecher, S., Stafford, M., & Wallace, E. (1980).

Screening Test of Adolescent Language. Seattle, WA:

University of Washington Press.

Ragosta, M. & Nelson, C., (1986). TOEFL and hearing impaired

students: A feasibility study. (Report No. ETS-RM-86-4). (ERIC

Document Reproduction Service No. Ed 269 422).

Semel, E., & Wiig, E. (1980). Clinical Evaluation of Language

Functions. Columbus, OH: Charles E. Merrill.

- Thorun, A. R. (1986). The Fullerton Language Test for Adolescents (2nd ed.). Palo Alto, California: Consulting Psychologist's Press, Incorporated.
- Wilcox, L. O. (1975). The prediction of academic success of undergraduate foreign students (Doctoral dissertation, University of Minnesota, 1974). Dissertation Abstracts International, 35, 6084B.

Appendix A

Review of Literature

New and more efficient ways of assessing language development and competency are continually being sought. One instrument that has been used to assess English language proficiency is the Test of English as a Foreign Language (TOEFL). The newly formed National Council on the Testing of English as a Foreign Language developed the TOEFL in 1963 for the purpose of assessing English language proficiency of foreign students applying for admission to United States colleges and universities. The TOEFL originally consisted of five sections but was revised to its present three section format in 1976. A new equivalent form of the TOEFL is developed for each consecutive administration. A TOEFL Committee of Examiners Policy Council, a panel of specialists in linguistics, language testing and teachers of English as a foreign or second language, continually reviews the TOEFL to ensure that the most current trends and methodologies in the field are reflected in the test's content (ETS, 1987). Individual test questions are reviewed internally by the Educational Testing Services for racial and cultural bias before inclusion in a test (ETS, 1987).

Colleges and universities throughout the United States and Canada currently use the TOEFL as a screening instrument to measure English language proficiency of foreign students seeking admission (Ayers & Peters, 1977; Bostic, 1981; Duran, Canale, Penfield, Stansfield, &

Liskin-Gasparro, 1985; ETS, 1987). One study, conducted at North Texas State University, found the TOEFL was the most highly attended to variable considered by administrators when considering foreign students for graduate admission (Gharavi, 1977). The TOEFL Program publishes official preparation and study materials. These materials are developed to aid students in understanding the format of the test and the special linguistic skills measured by the TOEFL (ETS, 1987). One of these aids, the TOEFL Sample Test (ETS, 1985), contains actual questions used in previous TOEFL administrations. Sold by ETS, the TOEFL Sample Test is a public domain, representational short form of the TOEFL, and is approximately one-half the length.

The Educational Testing Service (ETS, 1987) estimates the developmental difficulty of the TOEFL to be at the 11th grade level. The test evaluates receptive language ability in the areas of grammar, vocabulary, and usage (Duran et al., 1985). The currently administered three section TOEFL is presented in multiple choice format. The three sections are timed and assess receptive abilities in listening comprehension, structure and written expression, and vocabulary and reading comprehension (Alderman, 1982; Clark, 1977; ETS, 1987).

Because the TOEFL is so widely used as a method of screening foreign students, it has been the subject of much research directed

toward groups for whom English is a second language. Several studies have examined the predictive validity of the TOEFL.

Predictive Validity

Komvichayungyuen (1977) examined the relation between English language proficiency as rated by university faculty and the TOEFL. A cross sectional sample of 32 male and 25 female foreign speaking graduate students at Florida State University were included in the study. Each subject had been administered the five section TOEFL prior to admission to the university. Each subject was rated by one or more faculty members in the areas of writing, reading and vocabulary, listening comprehension, and speaking. Ratings were averaged across judges and area scores were totaled to express an over all English proficiency rating. This rating was compared to each subjects' TOEFL scores. Komvichayungyuen found a significant relationship between the two scores indicating the TOEFL to be a good predictor of the English language proficiency of foreign speaking students in real academic settings.

Research has shown the TOEFL to be inconsistent in its ability to predict first year academic success in foreign speaking students.

Wilcox (1975) examined the relationship between the TOEFL and first semester grade point average (GPA), both for a group of 99 students from Hong Kong, and a group of 84 Vietnamese students. Wilcox found a high degree of correlation between GPA and TOEFL

scores for the Vietnamese group and virtually no correlation for the group of Hong Kong students. Odunzes' study (1982) found no significant correlation between first year GPA and TOEFL scores for a group 260 Nigerian students. Andalib (1976) also found the TOEFL did not accurately predict GPA for a group of 126 Iranian students. Conversely, Abadzi (1976) and Ayers & Peters (1977) found TOEFL scores to be reasonable predictors of success for Asian students.

Concurrent Validity

Other studies have focused on the concurrent validity of the TOEFL. Dizney (1965) found TOEFL scores correlated highly with scores obtained on the Michigan Test of English Language Proficiency for a group of 20 foreign students. Alderman (1982) discovered a high degree of correlation between the TOEFL and the Test of English as a Second Language Achievement Test in a group of 411 Puerto Rican students from public and private schools. Angelis, Swinton, and Cowell (1979) reported that in foreign speaking applicants for university admissions the TOEFL did not correlate with verbal aptitude components of the Graduate Record Exam and the College Board Scholastic Aptitude Test.

Pack's 1971 study examined the relationship between the TOEFL and the Michigan Test of English Language Proficiency. TOEFL and Michigan test scores and freshman English grades were obtained for a group of 402 foreign students who attended the Church College of

Hawaii between September 1960 and February 1972. Pack found that although there was a significant correlation between the tests and grades in the first English class taken, there was only a moderate correlation between the TOEFL and Michigan scores.

Hosley (1978) conducted a study on 147 foreign students in the Center for English as a Second Language at the University of Arizona. The subjects were divided into five groups representing the five most common countries of origin. These groups included Mexico, Saudi Arabia, Libya, Venezuela and Japan. TOEFL scores for these subjects were obtained. An analysis of variance on the TOEFL scores demonstrated a significant country of origin effect. Specifically, Hosley reported TOEFL scores of Mexican subjects to be the highest and significantly different from the scores obtained by Saudi Arabian and Libyan subjects, which were the lowest. He also concluded that male and female differences were not significant.

Perkins and Pharisk (1980) focused on correlations between TOEFL scores and scores obtained on three standardized reading tests. Three groups totaling 110 subjects were included in the study. All subjects were enrolled in the Center for English as a Second Language at Southern Illinois University and were classified in the top two levels of English language proficiency at the Center. Each subject was administered one of three standardized reading tests approximately one week before the administration of the TOEFL. The reading tests

included the Nelson-Denny Reading Test, (developed for use in grades 9 through 16), the Iowa Silent Reading Test, (for grades 9 through 14), and the McGraw-Hill Basic Skills System Reading Test. Perkins and Pharisk found moderate correlations between TOEFL scores and standardized reading test scores on the Nelson-Denny Reading Test ($r = .49$) and the Iowa Silent Reading Test ($r = .46$). They also found a strong correlation between TOEFL scores and standardized reading test scores on the McGraw-Hill Basic Skills System Reading Test ($r = .91$).

Carlson, Bridgeman, Camp, and Waanders (1985) examined the relationship between TOEFL test scores and writing ability. They collected four writing samples from 638 students seeking admission to United States institutions. The sample represented three major foreign language groups including Arabic, Chinese, and Spanish. A small group of native English speakers were included in the sample. (Writing sample analysis for the native English group also was correlated to GRE scores and is not germane to this discussion.) Writing samples were analyzed and scored by 23 English as a Second Language (ESL) specialists and 23 English writing experts. Correlations for inter-rater reliability were found to be high. TOEFL scores were obtained for all subjects in the foreign sample. Correlations and factor analyses were performed between TOEFL and writing sample scores. These scores were highly correlated,

indicating that the writing sample analyses and the TOEFL are similar in their ability to assess English language proficiency.

Native Speakers and the TOEFL

The TOEFL is presented by its authors to be a robust measure of English proficiency, measuring the non-native prospective college student's ability in several areas of English language proficiency (ETS, 1987). The ETS contends that high scores on the TOEFL reflect a competency level in English that signifies readiness to cope with English language demands needed to pursue studies at United States colleges and universities. The TOEFL is not intended for use as a test of native English language proficiency. The ETS states that "the TOEFL test is recommended for students at the eleventh-grade level or above; the test content is considered inappropriate for younger students" (ETS, 1987, p.6). Although this is interpreted by many to represent a statement that the TOEFL reflect at least 11th grade or high school level equivalency of English proficiency within the areas tested, the actual developmental native English competency level demanded for success on the TOEFL never has been determined.

Few studies have looked at the usefulness of the TOEFL with adolescent and young adult native speakers of English. Clark (1977), administered the TOEFL to 88 native English speaking high school seniors prior to graduation. Assuming the TOEFL was developed to present language tasks similar to those encountered by foreigners in

the United States, Clark postulated that an educated English speaker would be capable of meeting or exceeding the test's standard of language performance. If the native speakers were found to have no difficulty with the test material, the assumption was that the TOEFL test questions were appropriate for determining English competency in non-native speakers. The results of the study found that the TOEFL, on the whole, was easy for this population. Total test scores indicated that the native English speaking group performed extremely well on the test, and the Clark data are the apparent basis for the ETS's statement that the TOEFL is recommended for students at the 11th grade level or above. Clark also conjectured that due to ceiling effects seen in these native English speaking graduating seniors, any use of the TOEFL to measure language proficiency among native speakers would be virtually useless (Clark, 1977).

Angoff and Sharon (1971) tested a group of 71 native English speaking college freshman and found the TOEFL to be relatively easy for this population. Angoff and Sharon concluded that the TOEFL was not an appropriate test of English for native English speaking students. Johnson's (1977) findings in his study of 173 American freshman at the University of Tennessee were similar to previous studies that examined the performance of native English speaking college students on the TOEFL. Johnson added the conclusion that

regional English dialect had little effect on the native English speaking subject's ability to perform on the TOEFL.

Few TOEFL studies have been conducted on handicapped populations. Ragosta and Nelson (1986) used the TOEFL to measure the language proficiency of profoundly hearing impaired native English language college students. The TOEFL was presented to 26 college students 18 of whom were deaf from birth and 24 being deaf by age 15. All were proficient in American Sign Language. Test instructions were spoken, and listening comprehension test items were played on a tape recorder. An interpreter simultaneously signed the same words. The students generally scored in the 2nd quartile, better than chance performance but considerably lower than foreign college students' scores. Ragosta and Nelson concluded that the TOEFL appeared to be a valid measure of relative English language proficiency in deaf college students. This study established the usefulness of the TOEFL in evaluating language proficiency in communication disordered native speaking populations.

Assessing language proficiency in adolescents

Experienced examiners have reported that assessment of adolescent language is a difficult task. This is in part due to noncompliant test taking behaviors often seen in this age group. On the basis of their clinical experience, Emerick and Haynes (1986) offered several reasons why adolescents often are difficult to test and resistant to

therapy. One obvious factor is the many pressures and changes experienced by adolescents as they are developing physically, sexually and emotionally. Secondly, they are struggling with a desire to be independent, yet not wanting to be singled out of the crowd in any way that would suggest they are different from their peers. Any or all of these factors may have a detrimental effect on the interaction between the examiner and adolescent.

Difficulties in assessment of this population also may be due in part to inadequacies in existing published assessment instruments.

Compton (1984) reviewed 75 formal and informal standardized tests used in special education. The purpose of this review was to provide basic information about the test's purpose, format, content, procedures, limitations and to aid in the interpretation of results. Compton pointed out that each test has limitations and may not be recommended for use with all populations. The study by Duran and colleagues (1985) completed an in depth item and format analysis of a TOEFL test that concluded the TOEFL to be an appropriate instrument for assessing English language proficiency. Duran et al. also pointed out that many of the major language tests used today are more reliable, objective, and administratively efficient than they are valid, acceptable and relevant.

Several American English language tests currently used to test native speaking adolescent language abilities are the Test of Adolescent

Language (TOAL) (Hammill, Brown, Larson, & Wiederholt, 1980), the Fullerton Language Test for Adolescents, 2nd Edition (FLTA) (Thorun, 1986), the Peabody Picture Vocabulary Test - Revised (PPVT-R) (Dunn, 1981), Clinical Evaluation of Language Functions (CELF) (Semel & Wiig, 1980), and the Screening Test of Adolescent Language (STAL) (Prather, Breecher, Stafford & Wallace, 1980).

Lieberman, Heffron, West, Hutchinson, and Swen (1987) presented three diagnostic adolescent language tests and one adolescent language screening test to a group of 30 sixth-grade students. This group included 9 males and 21 females. Over a two week period, each student was presented four language measures including the TOAL, FLTA, CELF, and the STAL. Presentation sequence was counterbalanced to control for order effect. Lieberman et al., reported that although the content and procedures of administering the TOAL, FLTA, CELF, and the STAL differ from one another, student performance correlations between them ranged from moderate (.651) between CELF and STAL to high (.846) between the FLTA and the TOLD. There were also no significant differences found for pass/fail rates on any of the diagnostic tests. Lieberman et al., concluded that there may be substantial areas of overlap or redundancy among the four measures. Therefore, none of the four adolescent language measures was recommended for use above the others.

The TOAL (Hammill, Brown, Larson & Wiederholt, 1980) assesses the adolescent's receptive and expressive vocabulary and grammar in written and spoken forms. The test administers eight subtests for the purpose of measuring interactions among receptive vocabulary, receptive grammar, speaking vocabulary, speaking grammar, reading vocabulary, reading grammar, writing vocabulary and writing grammar. Subtests are not timed but each requires approximately fifteen minutes to administer. Following TOAL standardization protocol, all subtests, with the exception of Speaking/Vocabulary and Speaking/Grammar, can be group administered. Raw scores and percent correct scores are obtained for each subtest. The TOAL scoring manual contains tables for conversion of raw scores to percentiles, stanines, and other standard scores.

The Fullerton Language Test for Adolescents (Thorun, 1986) consists of eight norm referenced subtests. Compton (1984) reviewed the first edition of the FLTA and stated that the eight subtests were designed to differentiate language impaired adolescents from normal language adolescents and to identify strengths and weaknesses in language comprehension and production of these individuals. Two subtests assess receptive language and six assess expressive language. The subtests are not sequence dependent.

Dunn (1981) designed the PPVT-R as a standardized receptive vocabulary assessment tool that assesses receptive knowledge of nouns, verbs, and adjectives. The test has two alternative forms and is easy to administer and score. Scoring procedures provide the examiner with an age score equivalent based on norms from age 2 years to 40 years.

References

- Abadzi, J. (1976). Evaluation of foreign students' admission procedures used at the University of Alabama (Doctoral dissertation, University of Alabama, 1975). Dissertation Abstracts International, 36, 7754A.
- Alderman, D. L. (1982). Language proficiency as a moderator variable in testing academic aptitude. Journal of Educational Psychology, 74, 580-587.
- American National Standards Institute (1970). Specifications for audiometers (ANSI S3.6-1969). New York: ANSI.
- Andalib, A. A. (1976). The academic success of undergraduate Iranian students in selected Texas Universities (Doctoral dissertation, East Texas State University, 1975). Dissertation Abstracts International, 36, 4881A.
- Angelis, P., Swinton, S., & Cowell, W. (1979). The performance of non-native speakers of English on TOEFL and verbal aptitude tests. (Report No. ETS 3) Princeton, NJ: Educational Testing Service.
- Angoff, W., & Sharon, A. (1971). A comparison of scores earned on the TOEFL by native American college students and foreign applicants to U.S. colleges. TESOL Quarterly, 5, 129-136.
- Ayers, J. B., & Peters, R. M. (1977). Predictive validity of the Test of English as a Foreign Language for Asian graduate students in

- engineering, chemistry, or mathematics. Educational and Psychological Measurement, 37, 461-463.
- Bostic, M. L. (1981). A correlational study of academic achievement and the Test of English as a Second (sic) Language (TOEFL) (Doctoral dissertation, University of Chicago, 1981). Dissertation Abstracts International, 42, 468A.
- Carlson, S. B., Bridgeman, B., Camp, R. & Waanders, J. (1985). Relationship of admissions test scores to writing performance of native and nonnative speakers of English. (Report No. ETS 19) Princeton, NJ: Educational Testing Service .
- Clark, J. L. D. (1977). The performance of native English speakers on the Test of English as a Foreign Language. (Report No. ETS 1) Princeton, NJ: Educational Testing Service.
- Compton, C. (1984). A guide to 75 tests for Special Education. Belmont, California: Pitman Learning.
- Dizney, H. (1965). Concurrent validity of the Test of English as Foreign Language for a group of foreign students at an American University. Educational and Psychological Measurement, 25, 1129-1131.
- Dunn, L. M. (1981). The Peabody Picture Vocabulary Test - Revised. Circle Pines, MN: American Guidance Service.
- Duran, R. P., Canale, M., Penfield, J., Stansfield, C. & Liskin-Gasparro, J. (1985). TOEFL from a communicative viewpoint

- on language proficiency: A working paper. (Report No. ETS 17). Princeton, NJ: Educational Testing Service.
- Educational Testing Service (ETS), (1987). TOEFL Test and Score Manual. Princeton, NJ: Author.
- Educational Testing Service (ETS) (1985). Test of English as a Foreign Language (TOEFL) Sample Test. Princeton, NJ: Author.
- Emerick, L. & Haynes, W. (1973). Diagnosis and evaluation in speech pathology (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Gharavi, E. (1977). Admission of foreign graduate students: An analysis of judgments by selected faculty and administrators at North Texas State University (Doctoral dissertation, North Texas State University, 1977). Dissertation Abstracts International, 38, 1148A.
- Hammill, D., Brown, V., Larson, S., & Wiederholt, J. (1980). Test of Adolescent Language. Austin, TX: Pro-Ed.
- Hosley, D. (1978). Performance difference of foreign students on the TOEFL. TESOL Quarterly, 12, 99-100.
- Johnson, D. (1977). The TOEFL and domestic students: Conclusively inappropriate. TESOL Quarterly, 11, 79-86.
- Komvichayungyuen, N. (1977). Test of English as a Foreign Language as a predictor of actual language proficiency.

- (Doctoral dissertation, Florida State University, 1977).
Dissertation Abstracts International, 38, 6666A.
- Lieberman, R. J., Heffron, A., West, S., Hutchinson, E., & Swen, T. (1987). A comparison of four adolescent language tests. Language, Speech, and Hearing in Schools, 18, 250-265.
- Odunze, O. J. (1982). Test of English as Foreign Language and first year G.P.A. of Nigerian students. (Doctoral dissertation, University of Houston, 1981). Dissertation Abstracts International, 42, 3419A-3420A.
- Pack, A. (1972). A comparison between TOEFL and Michigan Test scores and student success in (1) Freshman English and (2) Completing a college program. TESOL Reporter, 9, 1-7.
- Perkins, K., & Pharisk, K. (1980). TOEFL scores in relation to standardized reading tests. Research in Language Testing, 142-146.
- Prather, E., Breecher, S., Stafford, M., & Wallace, E. (1980). Screening Test of Adolescent Language. Seattle, WA: University of Washington Press.
- Ragosta, M. & Nelson, C., (1986). TOEFL and hearing impaired students: A feasibility study. (Report No. ETS-RM-86-4). (ERIC Document Reproduction Service No. Ed 269 422).
- Semel, E., & Wiig, E. (1980). Clinical Evaluation of Language Functions. Columbus, OH: Charles E. Merrill.

- Thorum, A. R. (1986). The Fullerton Language Test for Adolescents (2nd ed.). Palo Alto, California: Consulting Psychologist's Press, Incorporated.
- Wilcox, L. O. (1975). The prediction of academic success of undergraduate foreign students (Doctoral dissertation, University of Minnesota, 1974). Dissertation Abstracts International, 35, 6084B.

Appendix B

Subject Qualifications Background Questions

1. Student's name _____ Age: ____ Sex: M, F
2. Year in school: (10,11,12).
3. Is English your son/daughter's native language? (N, Y)
4. Does your son/daughter speak any languages other than English?
(N, Y)
5. Have any languages other than English been spoken in your home?
N,Y (If yes please specify. _____)
6. Has your son/daughter ever received speech or language therapy?
N,Y (If yes please elaborate. _____)
7. Has your son/daughter ever been diagnosed as having a hearing
loss? N, Y
(If yes please elaborate. _____)
8. Has your son/daughter ever been recommended for or received
special education to help him/her with a reading, math, or other learning
problem? (Resource room; learning disabilities, special ed., etc.)
N, Y

I hereby confirm that the above information is true and correct to the best of my knowledge.

Signature (Parent or Guardian)

Date

Appendix C

Informed consent document

**Brigham Young University
Research Participant
Informed Consent Document**

Participant: _____

Date of Birth: _____ Grade: _____

Parent or Guardian: _____

Address: _____

Telephone: _____

Your High School student is asked to become one of 60 local volunteer participants in a research study sponsored by the Communicative Sciences and Disorders area of the Department of Educational Psychology at Brigham Young University, Provo, Utah. The faculty director of this research is Dr. Laurence M. Hilton Ph.D., and the field investigator is Paul G. Osborn B.S., a graduate student completing his thesis requirement for the Master of Science degree in Speech-Language Pathology.

This research will study the performance of high school students on three standardized tests of English proficiency. The names of the tests to be administered are: The Peabody Picture Vocabulary Test - Revised (PPVT-R), The Test of

English as a Foreign Language Sample Test (TOEFL-ST), and The Fullerton Language Test for Adolescents (FLTA). You as parent or guardian must complete a short general background history questionnaire (read and sign the attached questionnaire) to determine whether or not your student meets minimal eligibility requirements for participation.

The purpose of the research is to see how well native English speaking high school students do on the Test of English as a Foreign Language Sample Test (a representative short for of the TOEFL, a test required of foreign students to enter U.S. colleges and universities) as compared to their performance on two tests of adolescent language widely used in Speech and Language clinics to describe normal language abilities.

Because these tests require careful listening, each student will be required to pass a simple hearing screening test before being permitted to participate. This will involve identification of "tones" within the normal range for speech, presented at a soft volume through earphones. The student will listen and raise his hand when he hears the tone.

Each participant in the study will be given one or three different measures of English language proficiency. The first, the PPVT-R, is a listening vocabulary test which requires the student to look at several pictures and select the most appropriate one for each vocabulary word presented. Students who do not receive an average to above-average score for their age on the PPVT-R will not take the TOEFL-ST or the FLTA. The next test, the TOEFL-ST, is a language evaluation containing three sections. The first section will require the student to listen to an audio recorded

passage and answer questions about it. The second requires the student to recognize standard written English grammar by completing a sentence. The third requires the student to read a passage and answer questions about it. The third measure, the FLTA has four vocabulary and four grammar sections. The entire procedure will take about two hours and will be completed in one day.

There are no known risks associated with these procedures. The confidentiality of each participant will be respected. The results of each student's performance will not be recorded in any school records. No names or other personal information about the study will be released except upon the condition of written request of the parent or guardian. Participants are volunteers and are free to withdraw at any time.

Participation is a volunteer service and no payment or monetary reward of any kind is possible or implied.

YES: I give my permission for _____

(Above named participant)

to participate in the above mentioned Brigham Young University research study. I confirm that I have read the preceding information and hereby give my informed consent for participation as described.

Signature (Parent or Guardian)

Date