

A Performance Comparison of Native and Non-native Speakers of English on an English Language Proficiency Test

Introduction

According to the U.S. Census Bureau, the population of the United States grew by approximately 12.5 percent from 1995 to 2000. Crawford (1997) reports that language diversity has increased dramatically throughout the nation. The composition of American society has changed substantially over the past five years, which has led to an increase in the number of language minorities.

Consequently, schools have increasingly larger groups made up of students whose native language is not English. This change has influenced schools to a great extent, especially in their educational testing programs. According to Geisinger and Carlson (1992), 15 to 20 percent of school-aged children speak a foreign language at home and do not speak English as their first language.

The No Child Left Behind Act of 2001 (NCLB) has focused attention on the appropriate assessment of English language learners (ELL students) in U.S. public schools. For teachers of ELL students, the goal of helping their students attain English language proficiency is inherently complex. During the language acquisition process, ELL students often achieve conversational fluency within one or two years, but their ability to reach grade-appropriate academic proficiency can take up to five years or longer (Case, 2003). Cummins (1997) identified this linguistic phenomenon as basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP). For these students, language can generally be divided into social language and academic language.

Federal legislation now prohibits schools from simply exempting students from testing and requires schools to provide appropriate test accommodations or alternate tests instead. For the more than four million ELL students in K–12 public education across the nation, the federal expectation is that they will be able to function in regular classrooms within three years. Therefore, English language proficiency measures must have a meaningful relationship with requirements of the classroom.

Using the appropriate type of test for a specific purpose is important to the validity and fairness of that test. Therefore, to validate any test, it is important to gather evidence. One

A Performance Comparison of Native and Non-native Speakers of English

method of gathering evidence to support the validity of an English language proficiency test is to give the test to a group of native speakers of English and a group of ELL students and compare their scores. These scores can also be used to validate “proficiency” so that there is a rational link between the English proficiency of ELL students and native English speakers.

Purpose of Study

The primary purpose of this study is to compare the performance of ELL students and native English speakers on three subtest scores of an English language proficiency test. The three subtest scores examined in this study are Listening, Writing Conventions, and Reading. The two research questions that were examined were:

1. Are there group differences between native English speakers and non-native speakers (ELL students) on English language proficiency test scores for three subtests?
2. Can group membership (Native versus Non-Native speakers of English) be reliably predicted from the total of the three subtest scores?

Theoretical Framework

Discriminant analysis (Hair, et al., 1998) was used to assign students into two groups (Native and Non-Native) based on the classification variables (Listening, Writing Conventions, and Reading). The classification was based on Bayesian posterior probability of group membership by assigning each individual to the group with the highest probability of group membership. The classification based on the Bayesian posterior probabilities made use of the classification function coefficients. The classification of an individual into one of the two groups was accomplished by computing a composite score for each of the two groups using the coefficients and the individual scores on the classification variables. Individuals were assigned to a group according to the composite scores. The classification procedure was then applied to a new sample for cross validation.

Method

This study used a sample of students taking the multiple-choice portion (Listening, Writing Conventions, and Reading subtests) of the Stanford English Language Proficiency (Stanford ELP) test in the fall of 2002 and the fall of 2003. Approximately 70 school districts from 20 states were included in this study. The multiple-choice part of Stanford ELP is untimed and it is group-administered. It is comprised of four different test levels: Primary, Elementary, Middle Grades, and High School. The Primary test level assesses kindergarten through Grade 2. However, for this study, kindergarten was not included. The Elementary test level assesses Grades 3 through 5. The Middle Grades test

A Performance Comparison of Native and Non-native Speakers of English

level assesses grades 6 through 8. The High School level assesses grades 9 through 12. The analyses for this study were grouped by test level and subtest. For Primary and Elementary, each subtest consisted of 20 items. For Middle Grades and High School, the Listening subtest consisted of 20 items and the Writing Conventions and Reading subtests consisted of 24 items each. Random samples of 200 students were selected for each group (Native and Non-native) and for each level. The mean scores obtained on the Listening, Writing Conventions, and Reading subtests were then used to identify any group differences between native and non-native speakers of English.

Results

To answer the first research question: Are there group differences between native English speakers and non-native speakers on English language proficiency test scores for three subtests? Analyses of variance (ANOVA) were conducted on each of the three subtests by the four levels. Table 1 presents, by test level, the number of examinees (*N*) who took each subtest, mean score earned, standard deviation (*SD*), *F* value, and probability ($Pr > F$).

Table 1. Performance Differences between Native and Non-native Speakers of English

	Native Speakers			Non-native Speakers			F Value	Pr>F
	N	Mean	SD	N	Mean	SD		
Listening								
Primary	200	17.92	1.63	200	17.43	1.60	9.07	0.003
Elementary	200	16.06	3.05	200	15.13	3.49	8.00	0.005
Middle Grades	200	17.14	2.21	200	14.82	3.71	58.00	0.0001
High School	200	16.75	1.99	200	13.43	2.94	175.14	0.0001
Writing Conventions								
Primary	200	14.79	3.51	200	13.18	4.37	16.49	0.0001
Elementary	200	17.22	3.02	200	14.48	4.01	59.91	0.0001
Middle Grades	200	20.63	3.08	200	15.46	4.56	176.36	0.0001
High School	200	20.12	2.73	200	15.80	4.50	134.78	0.0001
Reading								
Primary	200	14.70	3.86	200	12.76	4.62	20.75	0.0001
Elementary	200	16.56	3.38	200	13.35	4.33	68.32	0.0001
Middle Grades	200	20.88	3.13	200	15.64	4.91	161.94	0.0001
High School	200	21.62	3.19	200	17.07	4.79	125.11	0.0001

A Performance Comparison of Native and Non-native Speakers of English

Figure 1 illustrates the mean score difference between the Native and Non-native groups on the Listening subtest across the four test levels. Figure 2 depicts the mean score difference between the Native and Non-Native groups on the Writing Conventions subtest across the four test levels. Figure 3 presents the mean score difference between the groups on the Reading subtest. The results of the ANOVA showed that there were significant differences in the scores between the Native and Non-native groups on all four levels of the Listening, Writing Conventions, and Reading subtests. The native English speakers consistently scored higher than the non-native speakers of English.

Figure 1. Mean score difference between Native and Non-native groups

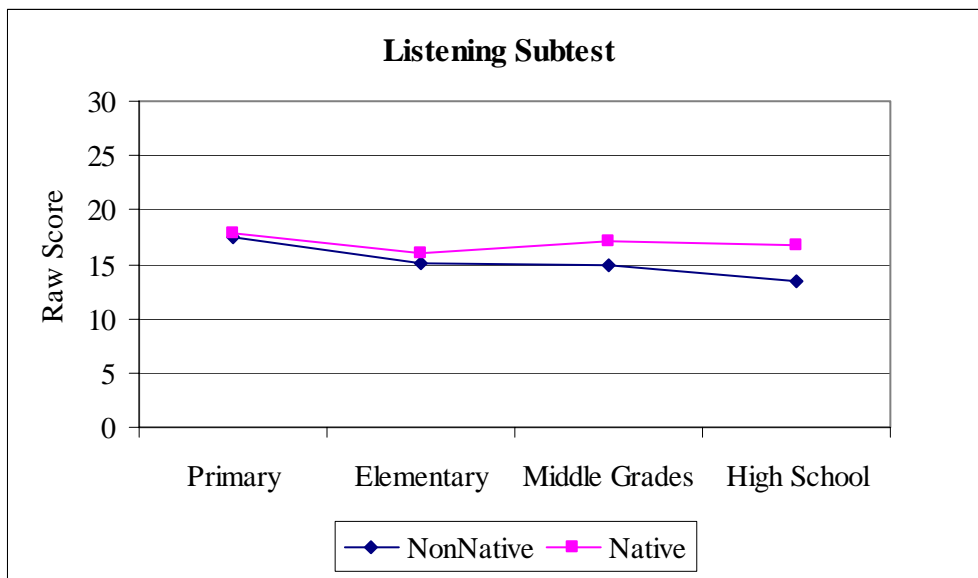
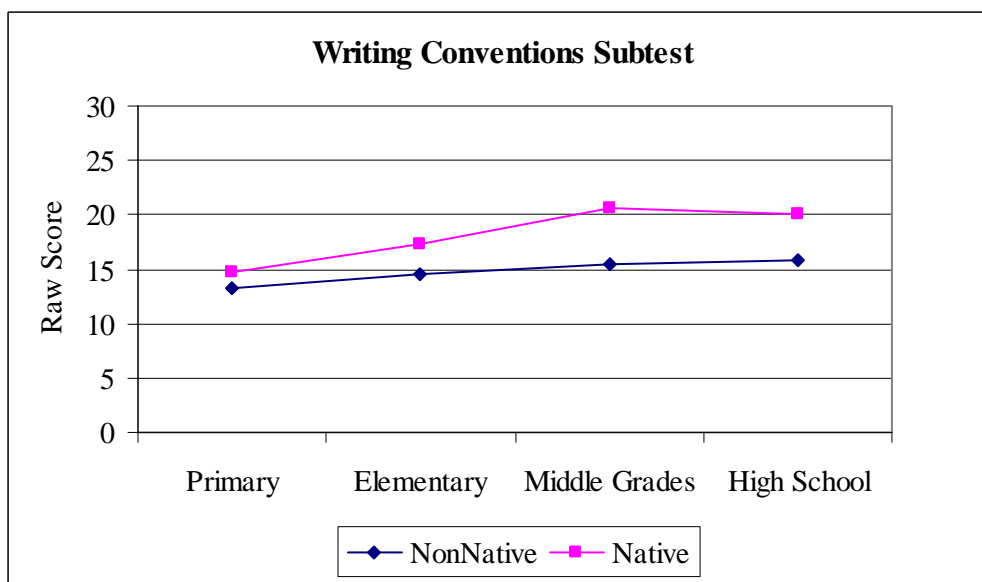
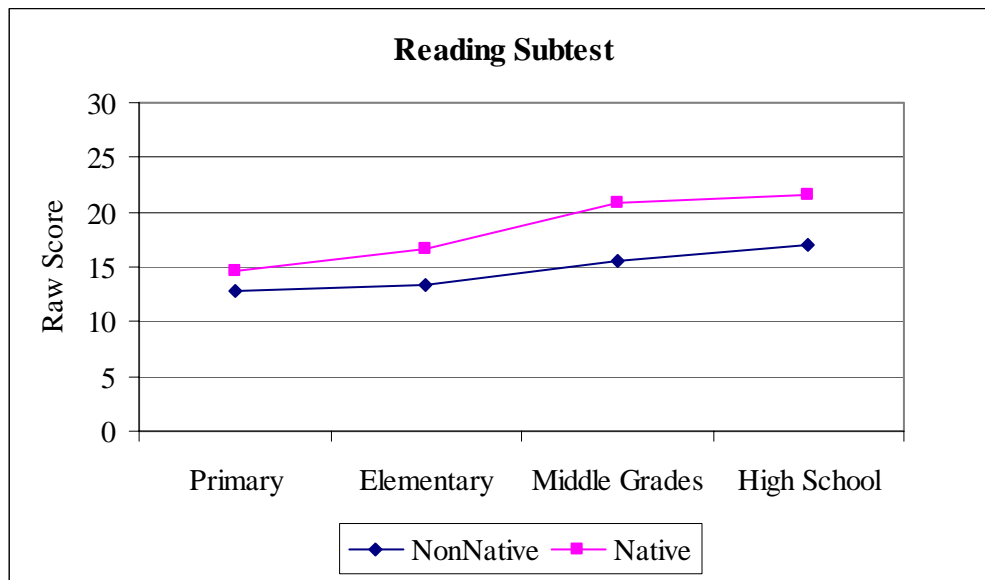


Figure 2. Mean score difference between Native and Non-native groups



A Performance Comparison of Native and Non-native Speakers of English

Figure 3. Mean score difference between Native and Non-native groups



For the Primary test level, the standardized function coefficient and correlation coefficients (see Table 2) revealed that the variables of Listening, Writing Conventions, and Reading were most associated with the function ordered by size of correlation. Fifty-six percent (56%) of the overall sample was correctly classified.

For the Elementary test level, the standardized function coefficient and correlation coefficients revealed that the variables of Listening, Writing Conventions, and Reading were most associated with the function ordered by size of correlation. Sixty-eight (68%) percent of the overall sample was correctly classified.

For the Middle Grades, the results revealed that the variables of Listening, Writing Conventions, and Reading were most associated with the function ordered by size of correlation. Eighty percent (80%) of the overall sample was correctly classified.

For the High School level, the results showed that the variables of Listening, Writing Conventions, and Reading were most associated with the function ordered by size of correlation. Eighty percent (80%) of the overall sample was correctly classified.

A Performance Comparison of Native and Non-native Speakers of English

Table 2. Correlation Coefficients and Standardized Function Coefficients

	Correlation Coefficient with Discrimination Function	Standardized Function Coefficient
Primary		
Listening	0.863	0.690
Writing Conventions	0.730	0.192
Reading	0.718	0.370
Elementary		
Listening	0.380	-0.470
Writing Conventions	0.770	0.453
Reading	0.926	0.897
Middle Grades		
Listening	0.580	-0.263
Writing Conventions	0.897	0.599
Reading	0.923	0.667
High School		
Listening	0.929	0.678
Writing Conventions	0.791	0.286
Reading	0.705	0.205

A Performance Comparison of Native and Non-native Speakers of English

Table 3. Classification Results

Actual Group	Number of Cases	Predicted Group	
		Native Speakers	Non-native Speakers
Primary			
Native Speakers	107	68(64%)	39(36%)
Non-native Speakers	102	53(52%)	49(48%)
Elementary			
Native Speakers	93	73(78%)	20(22%)
Non-native Speakers	87	38(44%)	49(56%)
Middle Grades			
Native Speakers	100	87(87%)	13(13%)
Non-native Speakers	102	27(27%)	75(74%)
High School			
Native Speakers	104	86(83%)	18(17%)
Non-native Speakers	111	26(23%)	85(77%)

Conclusion

From this study it can be concluded that for the Primary and Elementary test levels of Stanford ELP, the discriminant analysis was not very reliable—fifty-six and sixty-eight percent respectively. However, for the Middle Grades and High School levels, eighty percent of the sample was correctly classified. One of the reasons for the unstable classifications for the Primary and Elementary levels could be the fact that both native and non-native speakers of English have not attained sufficient knowledge in Writing Conventions and Reading subject areas.

References

- Case, B. (2003). *Accommodations on Stanford 10 for limited English proficient (LEP) students*. San Antonio, TX: Pearson Inc.
- Crawford, J. (1997, March). *Best evidence: Research foundations of the Bilingual Education Act*. Washington, D.C.: National Clearinghouse for Bilingual Education.
- Cummins, J. (1997). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Paper on Bilingualism*, No. 19, 121–129.
- Geisinger, K. F., & Carlson, J. F. (1992). *Practical assessment, research & evaluation: Assessing language-minority students*, 3(2).
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. Upper Saddle River, New Jersey: Prentice-Hall.

A Performance Comparison of Native and Non-native Speakers of English

- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1988). *Applied statistics for the behavioral sciences* (2nd ed.). Boston: Houghton Mifflin Co.
- Mertler, C. A. & Vannatta, R. A. (2001). *Advanced and multivariate statistical methods*. Los Angeles: Pyrczak.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tabachnick, B. G. & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York: Harper Collins.
- Tatsuoka, M. M. (1988). *Multivariate analysis: Techniques for educational and psychological research* (2nd ed.). New York: Macmillan.
- Young, M. J. and Yoon, B. (1998, April). *Estimating the consistency and accuracy of classifications in a standards-referenced assessment*. (National Center for Research on Evaluation, Standards, and Student Testing [CRESST] CSE Technical Report 475). Los Angeles, CA: University of California, Los Angeles.

Additional copies of this and related documents are available from:
Pearson Inc.
19500 Bulverde Rd.
San Antonio, TX 78259
1-800-211-8378
1-877-576-1816 (fax)
<http://www.hemweb.com/library/researchreports/index.htm>