

# **An Examination of the Validity of English Language Achievement Test Scores in a LEP Student Population**

Rodolfo Abella and Joanne Urrutia  
Miami-Dade County Public Schools, USA

**Abstract:** Approximately 1,700 limited English proficient (LEP) and recently exited LEP students in grades 4 and 10 were tested using both an English and a comparable Spanish language achievement test. Many LEP and former LEP students performed better in math taking the test in Spanish compared to taking it in English.

Educational accountability is increasingly demanded at a federal, state and local level. The emphasis on standards-based education requires that the public be informed on how students are performing in relation to these standards. In recent years a series of federal and state policies have mandated the testing of limited English proficient (LEP) students. More specifically, the No Child Left Behind Act of 2001 contains requirements directing states to implement assessment systems to annually measure the academic skills of all students, including LEP students.

There has been a great amount of debate about the pros and cons of testing LEP students. Some educators (Thurlow & Liu, 2001) consider it necessary, for accountability reasons, to include LEP students in all testing. Others (Oakeley & Urrabazo, 2001) think LEP students should be exempt from standardized testing programs because they lack English language skills and/or the cultural background to fully assimilate test questions. That is, existing evidence suggests that LEP students produce invalid test results (Abedi, Leon & Mirocha, 2001). Nevertheless, important decisions are currently being made on the basis of achievement test results, affecting individuals and organizations alike. In some states, students may graduate, teachers may get bonuses and schools may close on the basis of standardized test results. Consequently, it is crucial that academic progress be accurately assessed, including that of LEP students. The present study will address the issue of the test validity among LEP students by examining two factors that have been shown to affect LEP student performance on English language achievement tests: English language proficiency and home language literacy.

## **Method**

### *Sample*

A sample of 4<sup>th</sup> and 10<sup>th</sup> grade Miami-Dade County Public Schools students classified as Hispanic in district records was selected. Students were selected from those two specific grades to represent students in the primary and secondary grades, respectively.

Schools were selected districtwide on the basis of their free/reduced lunch ratios, a measure that reflects the schools or neighborhood's socioeconomic status. All district schools were divided into quartiles on that basis and equal numbers of schools were selected from each quartile. Within each quartile, schools with the highest number of LEP students were selected. Therefore, five schools per quartile were chosen at the elementary level (20 schools) and four per quartile were chosen at the senior high level (16 schools).

Both LEP and recently exited LEP students were included in the sample. LEP students were both in the beginning stages (these students are classified by the district as ESOL Levels I

and II) and advanced stages (ESOL Levels III and IV) of English language acquisition. All LEP students in the sampled elementary schools were selected to participate in the study. At the senior high level, 2 to 3 intact classrooms containing mostly LEP students, those taking a required language ESOL (English for Speakers of Other Languages) course, were selected randomly for inclusion in the sample. To select students who recently had exited the LEP program, four elementary and four senior high schools with the largest concentrations of LEP students were chosen from the sampled schools. All recently exited students in those schools were included in the sample. Students in this group are former LEP students who exited the ESOL program within the past two years and are now enrolled in regular curriculum courses. These students will be referred to in this paper as *Recently exited*.

As a result of this selection process, 712 LEP students and 223 recently exited students were tested in the 4<sup>th</sup> grade. Similarly, 581 LEP students and 163 recently exited students were tested in the 10<sup>th</sup> grade. The percentage of sampled LEP students receiving free/reduced lunch in the 4<sup>th</sup> (84%) and the 10<sup>th</sup> grade (62%) was the same as that of similar grade LEP students in the District. Overall, approximately one-fourth of all LEP students in the District were selected for testing in the 4<sup>th</sup> grade (25%) and 10<sup>th</sup> grade (22%).

#### *Test Materials and Data Collection Procedure*

Sample students were assessed in reading and mathematics using two comparable achievement tests, one in English (Stanford-9) and one in Spanish (Aprenda-2). Florida uses the *Stanford Achievement Test 9<sup>th</sup> Edition* as part of its statewide assessment program administered annually to students in grades 3-10. Sample students completed the Stanford in March 2002. In April 2002, sample students were assessed with the *Aprenda: La Prueba de Logros en Espanol, 2<sup>nd</sup> Edicion*. Both the Stanford and the Aprenda are established tests of academic achievement published by the same company, Harcourt-Brace. The Aprenda was modeled after the Stanford. According to the Aprenda technical manual, "...Aprenda 2 was planned to mirror the content and processes measured by the *Stanford Achievement Test, 9<sup>th</sup> Edition* (Harcourt-Brace, 1998, p. 8).

A key component of the present study is the comparison of student performance in mathematics across the two tests. Mathematics assessment used in this research focuses on the problem solving subtest at the 4<sup>th</sup> grade level and the total mathematics subtest in 10<sup>th</sup> grade. Each of these mathematics assessments consists of 48 multiple choice items. It should be noted that these items cover the same content in the two languages. According to the Aprenda technical manual the mathematics subtest A "was taken directly from Stanford-9 in order to help provide a statistical link between the two batteries" (Harcourt-Brace, 1998, p. 10).

## **Results**

### *Level of English Language Proficiency*

Recent studies have shown a limited relationship between language proficiency and English language achievement test scores among LEP students (Abedi, 2001; Stevens, Butler, & Castellon-Wellington, 2000). The analyses that follow further examine these variables, looking at the relationship between English language proficiency and performance on an English language achievement test (Stanford-9). In particular, the test performance of beginning and advanced LEP students is compared to that of students who recently, i.e., within the past two years, exited the ESOL program and also to that of Hispanic students in the district enrolled in the regular curriculum. In a second set of analyses, the validity of these scores is considered.

First, a series of one-way ANOVAs compared the mean scale scores of students across levels of language proficiency for each subtest and grade. The results show significant differences in Stanford mean scale scores among students on the basis of language proficiency, i.e., beginning, advanced, exited, and regular curriculum in both reading and math in grades 4 and 10. Generally, scores tend to increase the greater the level of English language proficiency (Table 1).

Table 1

*Mean Scale Scores by Language Proficiency Level on Stanford 9 Reading Comprehension and Mathematics; Grades 4, 10*

English Language Proficiency	GRADE 4			GRADE 10		
	n	Reading	Math	n	Reading	Math
Beginning	486	590	593	300	645	685
Advanced	226	619	616	281	671	702
Recent Exit	223	635	631	163	675	695
District Hispanic	437	643	635	363	696	706
F		219.09	137.47		188.70	38.03

*Note.* All F significant at  $p < .001$ . In order to have equivalent sample sizes, a random sample of 800 students was selected from grades 4 and 10 to represent the District's Hispanic student population in the regular curriculum. Mean scale scores for the random samples and the corresponding populations are identical.

A second set of analyses was conducted to verify whether the Stanford test results for LEP and recently exited LEP students were true measures of their content area knowledge or, instead, a function of language acting as a confounding variable. The performance of students in both the English (Stanford-9) and Spanish (Aprenda-2) versions of the mathematics section of the test was examined. Both the Spanish and the English language versions of the Problem Solving (Gr. 4) and Mathematics (Gr. 10) component of the tests contain 48 items of a comparable nature. A series of t tests contrasted the mean raw scores on the two tests.

The number of items answered correctly on the Aprenda math test was subtracted from the number of items answered correctly on the Stanford to create a difference score, which is displayed in Table 2. A positive score in the Math Diff. Score column in Table 2 indicates that the students, on average, answered a greater number of items correctly in the English language version of the mathematics test. A negative score indicates the reverse, students as a group answered more math items correctly in Spanish. It is assumed that the test version, i.e., language, which produces the greater number of correct answers is the more accurate, i.e., valid, assessment tool.

The results show that beginning LEP students in the 4<sup>th</sup> grade and all students in the 10<sup>th</sup> grade, both LEP and recently exited LEP, perform significantly better in the Spanish language test than in the English test. That is, they answered, on average, more items correctly on the Spanish version of the math test. Recently exited students in 4<sup>th</sup> grade performed significantly better in the English language test. Although the difference was not significant, 4<sup>th</sup> graders with advanced English language skills perform better on the Spanish language test (Table 2).

Table 2

*Mean Scale Scores and Mean Differences in Mathematics Raw Scores: Stanford 9 – Aprenda 2, by Language Proficiency Level, Grades 4 and 10*

English Language Proficiency	GRADE 4			GRADE 10		
	Math Diff. Score	t	d	Math Diff. Score	t	d
Beginning	-5.0	-18.83*	.59	-3.4	-8.11*	.45
Advanced	-0.4	1.25	.04	-2.2	-5.31*	.29
Recent Exit	1.8	4.76*	.24	-3.2	-6.58*	.47

*Note.* All t marked (\*) significant at  $p < .001$  level (paired t test, 2 tailed). d is Cohen's effect size.

### *Home Language Literacy*

Studies have shown that home language literacy is related to performance on standardized achievement tests (Hafner, 2001). To test this assumption, students were divided into three groups of equivalent size, comprising high, medium and low home language literacy, according to their scores on the reading comprehension section of the Aprenda.

A one-way ANOVA was used to contrast the mean math scale scores of the three home language literacy groups at each level of English language proficiency and grade (Table 3). The results show that English language math achievement test results vary on the basis of home language literacy. Students score significantly higher in math the higher their home language literacy skills. This is true in 4<sup>th</sup> and 10<sup>th</sup> grade and at all levels of English language proficiency.

The relationship of home language literacy to test validity was also examined. A series of one-way ANOVAs compared the math difference scores of students across levels of home language literacy for each level of English language proficiency and grade. The results show that students tend to perform significantly better in the Spanish math test, relative to the English math test, the higher their home language literacy. This is particularly true among students with advanced English language skills and among recently exited LEP students (Table 4).

A look at the math difference scores indicate that, at the fourth grade level, LEP students for the most part performed better in the Spanish version of the math test. The performance of recently exited students was better in English except for the high home language literacy group which performed about equally well in both languages. At the tenth grade level, students

performed better in Spanish versions of the test, across all levels of both home language literacy and language proficiency (Table 4).

Table 3

*Mean Scale Scores on Stanford 9 Mathematics by English Language Proficiency Level and Home Language Literacy Level, Grades 4 and 10*

English Language Proficiency	GRADE 4				GRADE 10			
	Low Home Lang.	Med. Home Lang.	High Home Lang.	F	Low Home Lang.	Med. Home Lang.	High Home Lang.	F
Beginning	570	594	614	91.82	674	681	699	27.49
Advanced	590	619	643	57.41	688	701	717	28.51
Recent Exit	616	631	651	26.16	686	691	706	9.28

*Note.* All F significant at  $p < .001$ .

Table 4

*Mean Differences in Mathematics Raw Scores: Stanford 9 - Aprenda 2, by Language Proficiency Level and Home Language Literacy, Grades 4 and 10*

English Language Proficiency	GRADE 4				GRADE 10			
	Low Home Lang.	Med. Home Lang.	High Home Lang.	F	Low Home Lang.	Med. Home Lang.	High Home Lang.	F
Beginning	-4.2	-5.2	-5.7	2.27	-2.4	-5.1	-2.6	4.09*
Advanced	1.6	-1.5	-1.7	9.49*	-0.9	-2.6	-3.5	3.22*
Recent Exit	1.8	1.6	-0.2	12.06*	-0.9	-4.2	-5.4	8.00*

*Note.* All F marked by (\*) significant at  $p < .05$  level.

In summary, students with high home language literacy skills are most likely to be negatively affected by participation in English language achievement tests

### Conclusions

The present results suggest that English language achievement tests are, for the most part, not a valid measure of content area knowledge in LEP students or in secondary students who have recently become language proficient. English language achievement tests appear to be particularly unable to accurately measure the content area skills of secondary students and of students with strong home language literacy backgrounds.

Important decisions made on the basis of achievement test results, such as promotions and high school graduation, should be reconsidered when LEP students and recently exited students are involved. The present findings indicate that the brightest immigrant students are the ones most likely to be penalized by decisions made on the basis of achievement testing. Researchers should attempt to replicate these findings using other populations and languages. If the present findings prove robust, alternative assessment and accountability procedures for the LEP and the recently exited student population should be created.

### References

- Abedi, J. (2001, April). *Validity considerations in the assessment of LEP students using standardized achievement tests*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Abedi, J., Leon, S., & Mirocha, J. (2001, April). *Validity of standardized achievement tests for English language learners*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Hafner, A. L. (2001, April). *Evaluating the impact of test accommodations on test scores of LEP students & non-LEP students*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Harcourt Brace. (1998). *Apreda technical manual* (2<sup>nd</sup> ed.). San Antonio, TX: Author.
- Holmes, D., Hedlund, P., & Nickerson, B. (2000). *Accommodating English language learners in state and local assessments*. Washington, DC: National Clearinghouse for Bilingual Education.
- Oakeley, C. & Urrabazo, T. (2001, April). *New state LEP testing policy in Texas: Is it an appropriate accountability measure for recent ESL immigrants?* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Stevens, R. A., Butler, F. A., & Castellon-Wellington, M. (2000). *Academic language and content assessment: Measuring the progress of English Language Learners (ELLs)*. University of California, L.A., Center for the Study of Evaluation. (CSE Technical Report 552).
- Thurlow, M. & Liu, K. (2001). *State and district assessments as an avenue to equity and excellence for English language learners with disabilities* (LEP Projects Report 2). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved January 10, 2002, from the World Wide Web:  
<http://education.umn.edu/NCEO/OnlinePubs/LEP2.html>.