

Miami-Dade County Public Schools  
Office of Program Evaluation  
1500 Biscayne Boulevard  
Miami, Florida 33132

**EVALUATION OF THE  
EXTENDED FOREIGN LANGUAGE PROGRAM  
FINAL REPORT**

**December, 2007**

Principal Evaluator/Author: Alex Shneyderman, Ed.D.

## Table of Contents

|   |     |
|---|-----|
| List of Tables .....                            | ii  |
| List of Figures .....                           | iii |
| Executive Summary .....                         | iv  |
| Description of the Program .....                | 1   |
| Design of the Evaluation .....                  | 2   |
| Sampling .....                                  | 3   |
| Instruments .....                               | 7   |
| Data Analysis .....                             | 10  |
| Results of the Evaluation .....                 | 10  |
| Evaluation Question 1 .....                     | 10  |
| Evaluation Question 2 .....                     | 13  |
| Evaluation Question 3 .....                     | 17  |
| Conclusions and Recommendations .....           | 20  |
| References .....                                | 22  |
| Appendices .....                                | 23  |
| Appendix A. List of Participating Schools ..... | 23  |
| Appendix B. 2004 Teacher Questionnaire .....    | 24  |
| Appendix C. 2007 Teacher Questionnaire .....    | 26  |

## List of Tables

|   |    |
|---|----|
| Table 1. Demographic and Achievement Characteristics of Students in the Two Samples.....  | 5  |
| Table 2. Demographic Characteristics of Students in the Two Models of the Program .....   | 6  |
| Table 3. Reading Achievement Growth Parameters for Students in the Two Subsamples .....   | 16 |
| Table 4. Mathematics Achievement Growth Parameters for Students in the Two Subsamples.... | 17 |
| Table 5. Perceived Effectiveness of the EFL Program .....                                 | 18 |

## List of Figures

|  |    |
|--|----|
| Figure 1. Aprenda 2 Reading Comprehension Results for Students in the Two Models of the Program.....     | 11 |
| Figure 2. LAS-O Results for Students in the Two Models of the Program.....                               | 12 |
| Figure 3. LAS-O Percentages in Performance Categories for Students in the Two Models of the Program..... | 12 |
| Figure 4. FCAT NRT Reading Achievement Results for Students in the Two Samples .....                     | 14 |
| Figure 5. FCAT NRT Mathematics Achievement Results for Students in the Two Samples.....                  | 15 |

## Executive Summary

This report presents results of the longitudinal evaluation of the Extended Foreign Language (EFL) program conducted during the 2003-04 through 2006-07 period. The EFL program is a dual language program, in which students are instructed in both English and a language other than English (Spanish, Haitian Creole, or Italian). Most of the schools participating in the program implement the English/Spanish version of dual language instruction. The program is designed so that students who participate in it for one school year are automatically enrolled in the program the subsequent year. Two models of dual language instruction within the EFL program are implemented in the District's schools. Under the first model (Model A), students receive an average of one hour of daily Language Arts instruction in the program's second language. Under the second model (Model B), in addition to an hour of daily Language Arts instruction, students receive an average of thirty minutes of daily Science or Social Studies instruction in the program's target language.

The evaluation of the EFL program focused on schools that implemented the most common English/Spanish mode of the program. This evaluation investigated the effects of the program on students' acquisition of Spanish reading and oral skills. It also examined the program effect on student learning of basic academic subjects of reading and mathematics. In addition, the evaluation considered teachers' views on the implementation of the program and possible ways to make it more successful. A group of 33 elementary schools that implemented the program during the 2003-04 was selected. A group of 33 comparison elementary schools not participating in the EFL program was selected to match the program schools academically and demographically from all elementary schools in the district. Program and Comparison student samples of 418 students each were selected; students in these samples were matched academically and demographically. Students in the Program Sample were assessed annually on their reading comprehension skills in Spanish during the 2003-04 through 2006-07 evaluation period. In addition, program students were assessed on their Spanish oral skills at the beginning and the end of this period. 2004-2007 results of the Florida Comprehensive Assessment Test (FCAT) were used to make academic achievement comparisons in reading and mathematics. Teachers in the program (203 in all) responded to a teacher survey.

The evaluation questions and results are presented below.

**Evaluation Question 1: How well are students progressing toward attaining Spanish language proficiency?** Students in both models of the program exhibited similar reading comprehension results and similar rates of annual growth in reading comprehension in Spanish between 2004 and 2007. In terms of oral proficiency in Spanish, students from Spanish language backgrounds performed similarly regardless of the model of the program. Students from non-Spanish language backgrounds who participated in the Model B of the program had higher Spanish oral results in both 2004 and 2007 than did students who participated in the Model A of the program. Students in both models of the program showed similar rates of growth in Spanish oral skills during the 2004-2007 period regardless of Spanish as a native language status.

**Evaluation Question 2: How does the program affect student achievement on the Florida Comprehensive Assessment Test?** Students in the program performed at similar levels of

achievement on the reading and mathematics components of the FCAT NRT compared to demographically and academically comparable students. The annual rates of reading and mathematics achievement growth were also similar for the program and comparison students.

**Evaluation Question 3: What are the teachers' views on the implementation of the program?** The majority of teachers stated that the EFL program operated above average or very well in their schools. In addition, the majority of teachers surveyed in both years believed that the program was very effective in improving Spanish skills of students from Spanish language background. On the other hand, in both 2004 and 2007, only the minority of teachers indicated that the program was very effective in helping students from non-Spanish language background to acquire Spanish language skills. A number of teachers reported interruptions or reductions in instructional time in the program due to preparing students for the FCAT. A sizable proportion of respondents perceived the need for better instructional materials and improvement in school schedules to allow teachers for more planning time.

In accordance with these results, the following recommendations are offered:

1. Develop a plan to provide all students who participated in the EFL program throughout elementary school an opportunity to continue learning the second language in middle school.
2. Develop a plan to address language needs of students from non-Spanish language backgrounds within the EFL program with the goal of reducing the gap in Spanish language achievement between them and native Spanish speakers by the end of elementary school.
3. Develop a uniform district policy regarding a temporary halt or reduction in time of the Spanish language instruction within the EFL program to prepare students for the FCAT.
4. Develop a plan to address the needs of teachers in the EFL program for instructional and supplemental materials.

## Description of the Program

The Extended Foreign Language (EFL) program is a “school within a school” delivery system, in which students are instructed in both English and a language other than English (Spanish, Haitian Creole, or Italian). The program is designed so that students who participate in it for one school year are automatically enrolled in the program the subsequent year. Two models of dual language instruction within the EFL program are implemented in the District’s schools. According to the program’s design, under the first model (Model A), students receive an average of one hour of daily Language Arts instruction in the program’s second language. Under the second model (Model B), in addition to an hour of daily Language Arts instruction, students receive an average of thirty minutes of daily Science or Social Studies instruction in the program’s target language.

In the 2003-04 academic year (first year of the evaluation), there were 46 elementary, 9 middle, and 2 senior high schools that implemented the EFL program. (The K-8 Centers are included in the count for elementary schools.) The degree of implementation of the program varied among schools. Generally, when a school decides to implement the EFL program, students in only the lowest grade level of the school’s entire grade span have an opportunity to participate in the program during the first year of implementation. During the following year, the program expands to the lowest two grade levels and so on. By the 2003-04 school year, 19 of the 46 elementary schools, four of the nine middle schools, and both high schools implemented the program in the entire grade spans. (Of the 19 elementary schools, one school had the program functioning as an extended day program.)

As more schools elected to implement the EFL program, the program was gradually expanded. In February 2005, the School Board adopted a resolution calling for an expansion of the EFL program at the elementary level with the goal of having at least one program per elementary school within each feeder-pattern. This decision provided motivation to further expand the program. By the 2006-07 school year, the program had expanded to 69 elementary schools and K-8 centers, 14 middle schools, and 4 senior high schools.

The evaluation of the EFL program was conducted by the Office of Program Evaluation at the request of the Division of Bilingual Education and World Languages. The evaluation covers a period of four academic years beginning with 2003-04 and extending through the 2006-07 school year. The purpose of this report is to depict changes in the EFL students’ Spanish reading and oral proficiency that occurred during this four year period. A secondary objective is to contrast the reading and mathematics achievement of students participating in the EFL program with that of similar students not participating in the program. Finally, this evaluation will attempt to describe teachers’ views on the implementation and effectiveness of the program.

Specifically, the evaluation report attempts to answer the following evaluation questions:

**Evaluation Question 1:** *How well are students progressing toward attaining Spanish language proficiency?*

**Evaluation Question 2:** *How does the program affect student achievement on the Florida Comprehensive Assessment Test?*

**Evaluation Question 3:** *What are the teachers' views on the implementation of the program?*

## Design of the Evaluation

One evaluation question calls for determining the effect of the program on students' achievement on the Florida Comprehensive Assessment Test (FCAT). To address this question, two student samples were created. The two samples are described in the section that follows.

### *Sampling*

Two samples were selected; each was created following a two-stage process. The first stage was devoted to school selection, and the second stage to student selection.

#### *School Selection*

For the selection of program schools, only elementary schools that had implemented the program in the full range of grade levels, from the lowest grade level up to at least the second grade, with Spanish as the target language, were considered. This choice was made because Spanish was taught to more than 95% of all elementary students in the program. During the 2003-04 school year, there were 36 such schools. By the 2006-07 school year, two of these schools discontinued the program. As mentioned earlier, one elementary school had the program functioning in the extended day mode. The remaining 33 schools were included in the sample of program schools. The list of these 33 schools is given in Appendix A.

For each of the selected 33 EFL schools, a matching non-EFL elementary school similar in the composition of its student population and its average student achievement but without the EFL program was selected from all such schools in the District. Specifically, the following demographic variables were used in the matching process: the size of student population, percentages of students in various racial/ethnic groups, percentage of students eligible to receive a free or reduced price lunch (FRL), percentage of limited English proficient (LEP) students, percentage of gifted students, and percentage of students for whom Spanish was their native language. In addition, two academic variables were used in the matching process: median school percentiles on the reading and mathematics subtests of the norm-referenced component of the FCAT. The school matching process consisted of two steps. First, the propensity scores (Rosenbaum and Rubin, 1983) of having the EFL program in the school were computed for all elementary schools in the District using the 2003-04 values of all demographic and academic variables mentioned above as covariates in a binary logistic regression. Second, the multivariate and propensity matching algorithm (Sekhon, 2007) was used to select those schools in the District that best matched the 33 program schools on propensity scores as well as on all demographic and academic variables listed previously.

#### *Student Selection*

For the program schools, one second-grade program classroom was selected at random in each of the 33 schools. Further, only the students who began participating in the program in grades K or 1 were included in the sample. In addition, only those students who were in the program in both 2003-04 and 2006-07 school years were included in the sample of program students. This procedure resulted in selection of 418 students. All these students constituted the Program Sample.

For each of the 418 students in the Program Sample, his or her gender, race/ethnicity, the October 2005 free/reduced price lunch status, gifted classification status, limited English proficient status, and native language (Spanish or other) was identified. In addition, students' reading and mathematics scale scores on the 2004 Florida Comprehensive Assessment Test Norm-Referenced Component (FCAT NRT) were recorded. Then, using a multivariate matching algorithm, an attempt was made to find comparison students from matching schools who would match the 418 program students exactly on all demographic variables while minimizing the multivariate Mahalanobis distance between program and comparison students on the two academic variables.<sup>1</sup> This attempt was successful for 356 of the 418 students in the Program Sample. To find the remaining 62 matches, a requirement that the comparison students come from the matching school was eliminated. That is, the remaining 62 comparison students were selected from the pool of 33 comparison schools so as to exactly match the 62 program students on all demographic variables mentioned earlier and to be as close as possible on the two academic variables. In the end, 418 comparison students were selected as described above; these students constituted the Comparison Sample.

The demographic and academic achievement characteristics of the two samples are shown in Table 1. In this table, school-level characteristics reflect the two groups of 33 schools represented in each student sample, and student-level characteristics demonstrate the results of the individual student matches. The results of the matching process presented in Table 1 show that the two students groups were well matched in terms of the characteristics of the schools they attended. In addition, student groups were well matched at the individual level. That is, the demographic characteristics of the two groups were exactly the same and mean scale scores on the 2004 FCAT NRT Mathematics were virtually identical. The mean scale score on the 2004 FCAT NRT Reading for students in the Program Sample was about two scale score points higher than for students in the Comparison Group.

In addition to comparing the academic achievement of program students in the basic academic subjects of reading and mathematics to that of non-program students, the Spanish reading and oral skills of the two groups were also compared. Two different student groups were considered. The first group contained those students who participated in the program in the 18 schools that implemented the Model A of the EFL program, and the second groups contained students who were in the program in the 10 schools implementing the Model B of the program during the evaluation period. (Of the 33 original program schools, 5 schools changed the implemented model of the program during the 2006-07 academic year). Demographic characteristics of the two groups of student and schools are shown in Table 2.

A subsample consisting of 10 students from the Program Sample was selected from each school and their oral skills in Spanish were assessed. Two student groups were identified: non-native Spanish-speaking students and native Spanish-speaking students. Students within each language group were selected at random during the 2003-04 academic year; however, students whose native language was other than Spanish were selected first. Since most of the participants in the

---

<sup>1</sup> The usual Euclidean distance between two cases that have values on several variables is given by the square root of the sum of squared differences between the values of the variables. The Mahalanobis distance extends the ordinary Euclidean distance by taking into account correlations between variables.

EFL program were native Spanish-speaking students, this order of selection was used to pick a sufficient number of non-native Spanish speaking students, allowing for a reliable estimation of their average oral language skills. During the spring of 2007, the LAS-O was administered to those students in the originally selected group who were still participating in the EFL program. In all, 150 such students were assessed; however, the scores of three students were not used in the analyses because they were recorded as zeros, likely indicating invalid results. Of the remaining 147 students, 94 participated in Model A of the program and 53 participated in Model B.

Table 1

*Demographic and Achievement Characteristics of Students in the two Samples*

|  | Program Sample | Comparison Sample |
|--|----------------|-------------------|
| <b>School-Level Characteristics</b>          | n = 33         | n = 33            |
| Median Size of Student Population            | 926            | 915               |
| <i>Median Percentage of Students who are</i> |                |                   |
| Hispanic                                     | 92             | 89                |
| Eligible for the FRL program                 | 70             | 71                |
| LEP  | 33             | 32                |
| Of Spanish language background               | 82             | 79                |
| Gifted                                       | 5              | 6                 |
| <i>Median National Percentile on the</i>     |                |                   |
| 2004 FCAT NRT Reading                        | 58             | 58                |
| 2004 FCAT NRT Mathematics                    | 64             | 64                |
| <b>Student-Level Characteristics</b>         | n = 418        | n = 418           |
| <i>Percentage of Students who are</i>        |                |                   |
| Female                                       | 59             | 59                |
| Hispanic                                     | 89             | 89                |
| Eligible for the FRL program                 | 52             | 52                |
| LEP  | 1*             | 1*                |
| Of Spanish language background               | 78             | 78                |
| Gifted                                       | 22             | 22                |
| <i>Mean Scale Score on the</i>               |                |                   |
| 2004 FCAT NRT Reading                        | 622.8          | 620.5             |
| 2004 FCAT NRT Mathematics                    | 612.6          | 612.4             |

\* There were only three LEP students (as of October 2005) in each of the two samples.

Table 2

*Demographic Characteristics of Students in the two Models of the Program*

|  | Model A | Model B |
|--|---------|---------|
| <b>School-Level Characteristics</b>          | n = 18  | n = 10  |
| Median Size of Student Population            | 857     | 1025    |
| <i>Median Percentage of Students who are</i> |         |         |
| Hispanic                                     | 91      | 92      |
| Eligible for the FRL program                 | 64      | 67      |
| LEP  | 31      | 34      |
| Of Spanish language background               | 80      | 87      |
| Gifted                                       | 8       | 4       |
| <b>Student-Level Characteristics</b>         | n = 225 | n = 137 |
| <i>Percentage of Students who are</i>        |         |         |
| Female                                       | 54      | 62      |
| Hispanic                                     | 83      | 93      |
| Eligible for the FRL program                 | 46      | 64      |
| LEP  | 0*      | 1*      |
| Of Spanish language background               | 67      | 88      |
| Gifted                                       | 27      | 15      |

\* There was only one LEP student in the sample (as of October 2005) participating in the Model A, and 2 participating in Model B of the program.

## *Instruments*

Several assessment instruments were used to measure student academic achievement. Some of them were used to assess the attainment of the Spanish language proficiency, while others were used to measure academic achievement in core subject areas of reading and mathematics.

### *Spanish Language Proficiency Instruments*

To assess students' attainment of reading proficiency in Spanish, the reading comprehension component of the *Aprenda*, second edition (*Aprenda 2*), developed by the Harcourt Assessment, was used. This instrument is a nationally recognized academic achievement test in Spanish that was "... planned to mirror the content and processes measured by the *Stanford Achievement Test*, Ninth Edition (Stanford 9)" (Harcourt Brace Educational Measurement, 1998). *Aprenda 2* was administered annually to students in the Program Sample during the 2003-04 through 2006-07 period. An appropriate form of the *Aprenda 2* corresponding to the students' grade level was given to students as they progressed from the second to the fifth grade during that period.

To assess students' attainment of oral language skills in Spanish, the *Language Assessment Scales-Oral (LAS-O)*, developed by the CTB/McGraw-Hill, was used. This instrument was used to measure students' Spanish skills in the following modalities: vocabulary, listening comprehension, story retelling, and pronunciation. LAS-O was administered to selected students in each of the selected program schools in the fall of 2003 and in the spring of 2007.

### *Academic Achievement Instruments*

The reading and mathematics components of the FCAT NRT were used to make academic achievement comparisons between students in the Program and Comparison Samples. This test is constructed to have a vertical scale of measurement, which makes it appropriate for longitudinal analysis of achievement outcomes. The state uses a secure form of the *Stanford Achievement Test* as the FCAT NRT. In 2004, the ninth edition of the *Stanford Achievement Test* (Stanford 9) was used, whereas in 2004-2007 the tenth edition of the test (Stanford 10) was used. The change in the edition required a suitable approach to data analysis, which will be described later.

### *Survey Instruments*

To assess teachers' views on the implementation of the EFL program, two versions of the *Teacher Questionnaire* were developed in cooperation with the program staff. These instruments inquired about the teachers' views of the program, its implementation, program effects on students' acquisition or maintenance of the Spanish language, and other pertinent program aspects. The two versions of the Teacher Questionnaire had a core of common items but differed on several other items that were developed to address concerns regarding the functioning of the program in the beginning (2004) and the end (2007) of the evaluation period. The two versions of the Teacher Questionnaire are given in the Appendices B and C.

## *Data Analysis*

Statistical analyses suitable for addressing the evaluation questions posed and the data collected were implemented. A description of these analyses is presented separately for each evaluation question.

### *Evaluation Question 1: How Well are Students Progressing Toward Attaining Spanish Language Proficiency?*

To address this question, the results of the Aprenda 2 and LAS-O administrations were used. As mentioned earlier, the Aprenda 2 was annually administered to program students during the period of 2003-04 through 2006-07. Therefore, each student in the Program Sample was assessed up to four times during the evaluation period. That is, several Aprenda 2 test results were available for each student. In turn, program students were clustered within the 33 selected schools that implemented one of the two models of the program. Therefore, the data were hierarchical in nature with test results clustered within students who were clustered within schools. Consequently, a three-level hierarchical linear model (Raudenbush and Bryk, 2002) was used. In this model, the first level represented the linear growth of test results with time. The growth parameters (slope and intercept) from the first level of the model were entered as outcomes in the second (student) level with student characteristics used as predictors. The student-level variables considered were FRL status, gifted status, and Spanish as a native language status (as of October 2005), all of which were dichotomous. Of these variables, the FRL status can change with time. However, more than three-quarters of the students did not change their FRL status during the evaluation period. Consequently, instead of including this variable as the time-varying covariate in the temporal level of the HLM model, the FRL status was used at the student level. The third (school) level of the statistical model added school-level variables: the percentages of FRL students, LEP students, and students having Spanish as their native language, as well as the indicator variable for the model of the program implemented in a particular school. All of these student- and school-level variables are generally considered to be associated with student learning outcomes. Therefore, including them as covariates in the statistical model served the dual purpose of estimating the effects of these variables on the student achievement outcomes as well as adjusting the outcomes for student and school differences on those variables. Because 5 of the 33 schools represented in the Student Sample changed the implemented model during the 2006-07 school year, the Aprenda 2 results of students from only the remaining 28 schools (362 students in all) were considered for this analysis. All data analyses used in this evaluation used the conventional .05 level of statistical significance.

To assess students' learning of Spanish oral skills, the results of the two administrations of the LAS-O (in 2004 and 2007) were contrasted. Because the number of students with valid scores accessed via the LAS-O instrument in both years was rather small (147, or about 5 students per school on average) and because the students' oral skills were assessed only twice during the evaluation period, the reliable estimation of the oral proficiency growth trajectories through the HLM technique was not possible. Instead, a repeated measures general linear model (GLM) with the 2004 and 2007 test results as the within-subject factor was used to conduct the analysis. In this analysis, the model of the EFL program and students' Spanish as a native language status were used as the between-subjects factors. The student demographic variables (FRL status and

gifted status) were used as covariates in this analysis. Inclusion of these variables as covariates in the GLM analysis allowed for statistical adjustment for demographic differences between students in the two models of the EFL program.

*Evaluation Question 2: How Does the Program Affect Student Achievement on the Florida Comprehensive Assessment Test?*

To address this question, the three-level HLM model (time within students within schools) was used. This model was similar to the one described above regarding the analysis of the Aprenda 2 results: the same student- and school-level variables were considered. There was one difference, however. Because the state used Stanford 9 in 2004, but Stanford 10 in 2005-2007 as the FCAT NRT, the 2004 results were not used at the *time* level of the model, but were instead used as covariates at the *student* level. This treatment of the 2004 results allowed constructing the growth trajectories based on achievement outcomes measured by the same edition of the test, Stanford 10. In addition, it allowed for statistical adjustment of growth parameters for differences in the 2004 results. Separate HLM analyses were conducted for reading and mathematics components of the FCAT NRT. Reading and mathematics results of all students in the Program and Comparison Samples (418 students in each) from all selected schools (33 program and 33 comparison schools) were used in these analyses.

*Evaluation Question 3: What are the Teachers' Views on the Implementation of the Program?*

To address this question, teachers in the program were surveyed in 2004 and again in 2007. Two different versions of the Teacher Questionnaire were used. The results of the surveys were analyzed and tabulated.

## Results of the Evaluation

### *Evaluation Question 1: How Well are Students Progressing Toward Attaining Spanish Language Proficiency?*

As mentioned earlier, reading and oral proficiency in Spanish were assessed with different instruments and with different frequency. Consequently, different statistical methods of data analysis were used to address students' outcomes in the two language components. Below, the results of the analyses are presented separately for reading and oral proficiency in Spanish.

#### *Results of the Spanish Reading Proficiency Analysis*

The three-level hierarchical linear modeling technique was used to compare the Spanish reading comprehension results for students in the two models of the program. The separate growth trajectories were constructed for students in Model A and Model B of the program using the Aprenda 2 Spanish reading comprehension results in years 2004-2007. Student- and school-level demographic variables were used as covariates to adjust for differences in demographic characteristics between students and schools in the two models of the program. The results of fitting HLM equations indicated that the 2004 Aprenda 2 adjusted mean scale score of students in the Model B of the program was 618.8, which was 7.0 scale score points higher than that of students in Model A. The difference in the initial, 2004 mean scale scores was not statistically significant. Students in the Model B of the EFL program experienced a higher average achievement growth in Spanish reading comprehension compared with students in the Model A of the program, although the difference in adjusted mean rates of annual growth (1.0 scale score points per year) was not statistically significant. The 2007 adjusted mean scale scores on the reading comprehension of Aprenda 2 were 649.6 and 660.1 for students participating in Models A and B of the program respectively. The difference of 10.5 scale score points was not statistically significant. These results are shown graphically in Figure 1. In this figure, numbers in parentheses indicate national normative percentiles corresponding to adjusted mean scale scores. These figures indicate that students in both models of the program exhibited reading comprehension in Spanish that was higher than that of the average student in the national norm group during the 2004-2007 period. However, students in the EFL program had a lower relative standing in 2007 than in 2004 compared with the national norm group.

The results of the statistical analyses of Aprenda 2 results also indicated that students who had Spanish as the native language had the initial average advantage of 18.3 mean scale score points compared with other students regardless of the model of the program. Annual achievement growth rates of Spanish language background students were on average comparable to those of other students in the program. Consequently, students from the Spanish language background maintained their initial advantage over students from other language backgrounds throughout the 2004-2007 period.

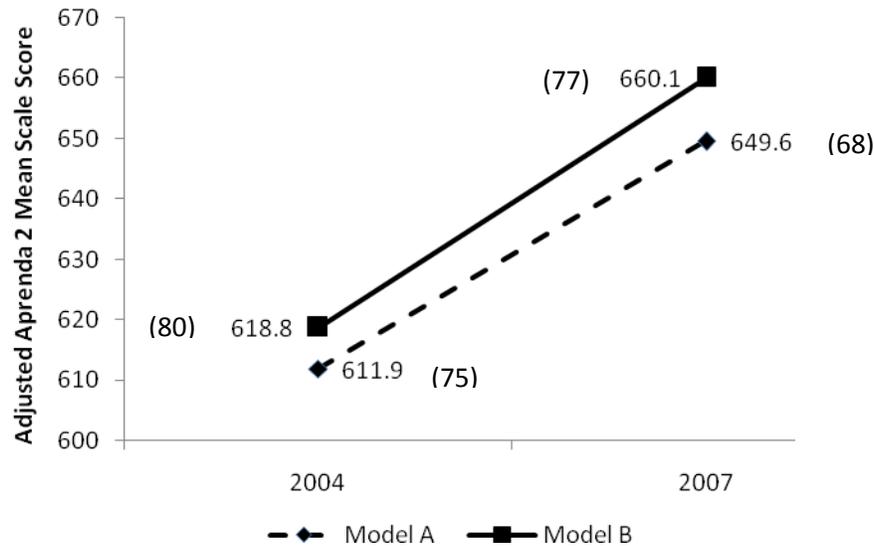


Figure 1

Aprenda 2 Reading Comprehension Results for Students in the Two Models of the Program

*Results of the Spanish Oral Proficiency Analysis*

The results of the GLM repeated measures analysis of the 2004 and 2007 LAS-O scores indicated that oral proficiency scores of students in both models of the program grew at similar rates between 2004 and 2007 (3.3 to 3.9 scale score points per year) regardless of the native language of students. That is, there was no statistically significant difference between the rates of growth of Spanish oral proficiency results on LAS-O between students in the two models of the program or having Spanish or another language as the native language during that period. However, there was a statistically significant simple main effect of the model for students who had a language other than Spanish as the native language. The LAS-O adjusted mean scale scores of such students participating in Model B of the EFL program were significantly higher than those in Model A in both 2004 (70.0 vs. 57.4) and 2007 (80.4 vs. 67.3). The 2004 and 2007 adjusted mean scale scores on LAS-O are shown graphically in Figure 2.

The results of the LAS-O are reported in terms of scaled scores and achievement levels ranging from 1 to 5. Level 1 is designated for students who showed low levels of Spanish oral skills. These students are classified by the test developers as non-Spanish speakers. Students who score within achievement levels 2-3 show higher levels of Spanish oral skills. These students are classified as limited Spanish speakers. Finally, students who score within achievement levels 4-5 are designated as fluent Spanish speakers. The percentages of students who scored within each proficiency classification in 2004 and 2007 are shown in Figure 3. These results are disaggregated by Spanish as a native language status. Because students who come from Spanish language background performed similarly regardless of the model of the program, their LAS-O results are not disaggregated by model.

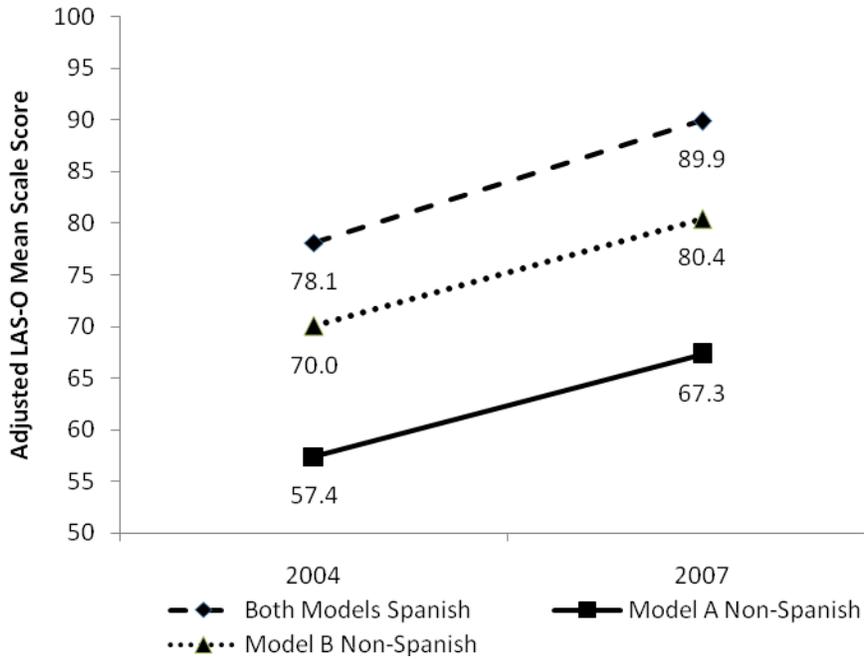


Figure 2

LAS-O Results for Students in the Two Models of the Program

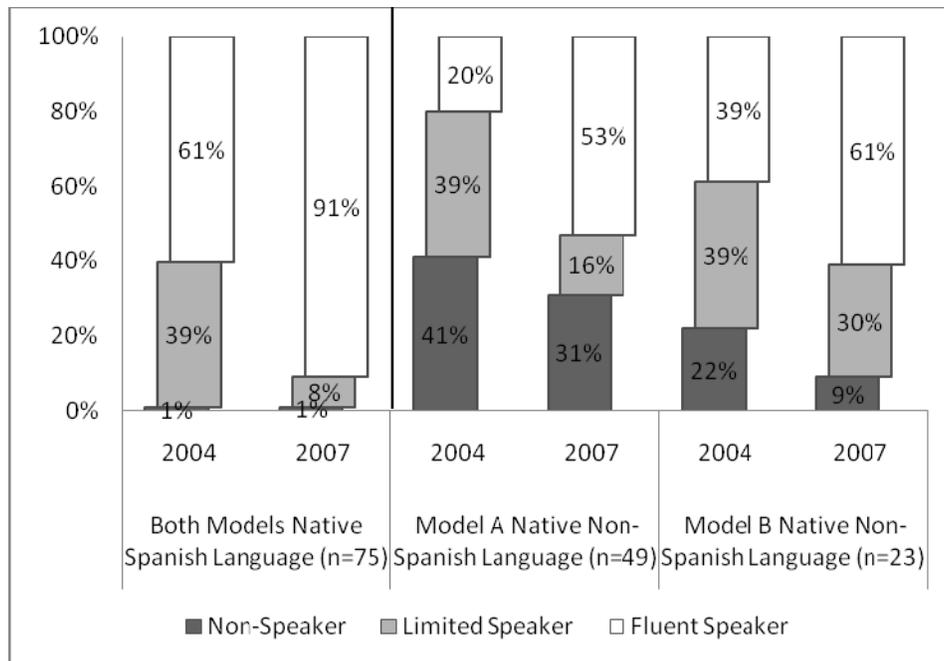


Figure 3

LAS-O Percentages in Performance Categories for Students in the Two Models of the Program

Figure 3 shows that students in both models of the program made improvement in their Spanish oral skills between 2004 and 2007. For students who had Spanish as the native language, the percentage of those who performed in the Fluent Spanish Speaker category on the LAS-O increased from 61% to 91% during that period. For students who had a language other than Spanish as their native language and participated in Model A of the program, the proportion of students scoring in the Fluent Spanish Speaker category increased from one-fifth to more than one-half during the same period. For students who had a language other than Spanish as their native language and were in Model B, that proportion increased from 39% to 61% during the 2004-2007 period.

*Summary for Evaluation Question 1.* Students in both models of the program exhibited similar rates of annual growth in reading comprehension in Spanish between 2004 and 2007. Those in Model B of the program had an initial advantage of seven scale score points in the Spanish reading comprehension results compared to students in Model A of the program. They were able to increase that advantage to 10.5 scale score points by 2007, although the differences in the adjusted mean reading comprehension scale scores between students in the two models of the program were not statistically significant in both 2004 and 2007. Students who had Spanish as the native language had an initial advantage of 18.3 scale score points compared with students who had another language as their native language; that advantage was maintained during the 2004-2007 period.

In terms of oral proficiency in Spanish, students from the Spanish language background performed similarly regardless of the model of the program. On the other hand, the model of the program had a differential impact on the acquisition of oral language skills in Spanish for students who had a language other than Spanish as their native language. Those students who participated in Model B of the program had a statistically significant higher mean scale score on the LAS-O in both 2004 and 2007 than did students who participated in Model A of the program. Students showed similar rates of growth in Spanish oral skills during the 2004-2007 period regardless of the model of the program or student language background. In addition, the proportions of such students who scored as Fluent Spanish Speakers grew during the 2004-2007 period for students in both models of the EFL program; these proportions were higher for students in Model B in both 2004 and 2007.

*Evaluation Question 2: How Does the Program Affect Student Achievement on Florida Comprehensive Assessment Test?*

The three-level hierarchical linear model (HLM) was used to construct linear growth trajectories of student achievement on the FCAT NRT during the three-year period from 2005 to 2007. Separate analyses were carried out for reading and mathematics achievement results. The 2004 student achievement results were used as covariates in these analyses. In addition, student- and school-level demographic variables were used as covariates to adjust the achievement growth parameters for differences between students in the Program and Comparison Samples. The results of the analyses are presented separately for reading and mathematics student achievement.

### Reading Achievement Results

The results of the reading achievement analysis indicate that the adjusted mean scale scores on the reading component of the 2005 FCAT NRT were 644.8 for students in the Program Sample and 641.4 for their counterparts in the Comparison Sample. The difference of 3.4 scale score points was not statistically significant. The average adjusted rates of annual growth in reading achievement were 24.3 and 23.6 for students in the Program and Comparison Samples respectively. The difference of 0.7 scale score points per year in the amount of growth in reading achievement was not statistically significant. These results are presented graphically in Figure 4. In this figure, numbers in parentheses reflect national percentiles corresponding to the adjusted mean scale scores. This normative information indicates that students in both samples scored better than two-thirds of students in the national norm sample on the reading component of the 2005 FCAT NRT; by 2007 they improved their relative standing outperforming more than 85% of students in the national norm sample.

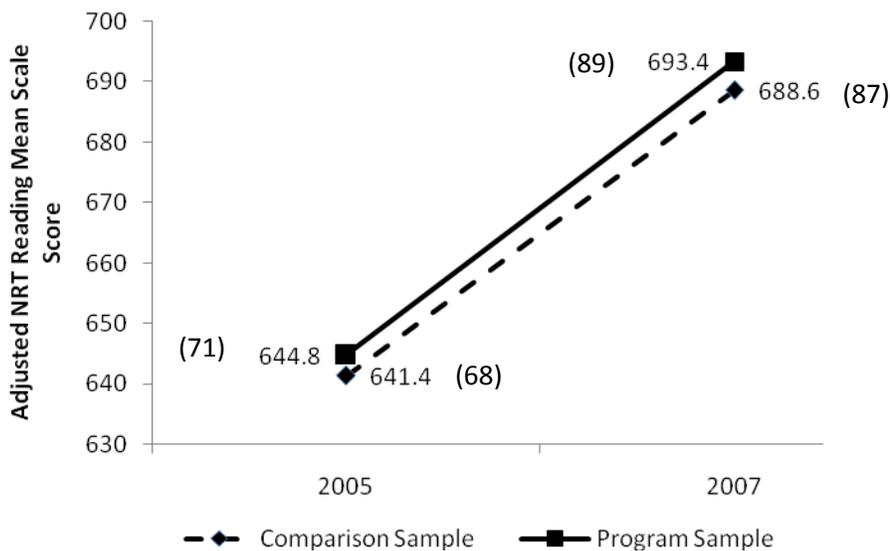


Figure 4

FCAT NRT Reading Achievement Results for Students in the Two Samples

### Mathematics Achievement Results

The analysis of the mathematics achievement results revealed a similar situation to the reading comprehension results when comparing the annual rates of growth in test performance. The adjusted mean scale score on the mathematics component of the 2005 FCAT NRT was 645.3, while that for students in the Comparison Sample was 652.3. The difference of seven scale score points was statistically significant. On the other hand, the annual rates of growth in mathematics achievement were similar for students in the two samples: the adjusted mean rate for students in the Program Sample was 16.5 scale score points per year, whereas that for students in the Comparison Sample was 15.3 scale score points per year. The difference of 1.2 scale score points per year in the amount of mathematics achievement growth was not statistically significant. These results are presented graphically in Figure 5. National normative information is shown in parentheses. It can be seen that in 2005 students in both samples performed better than three-fourths of students in the national norm sample on the mathematics component of the FCAT NRT. Students' relative standing improved by 2007: they outperformed more than 80% of students in the national norm sample.

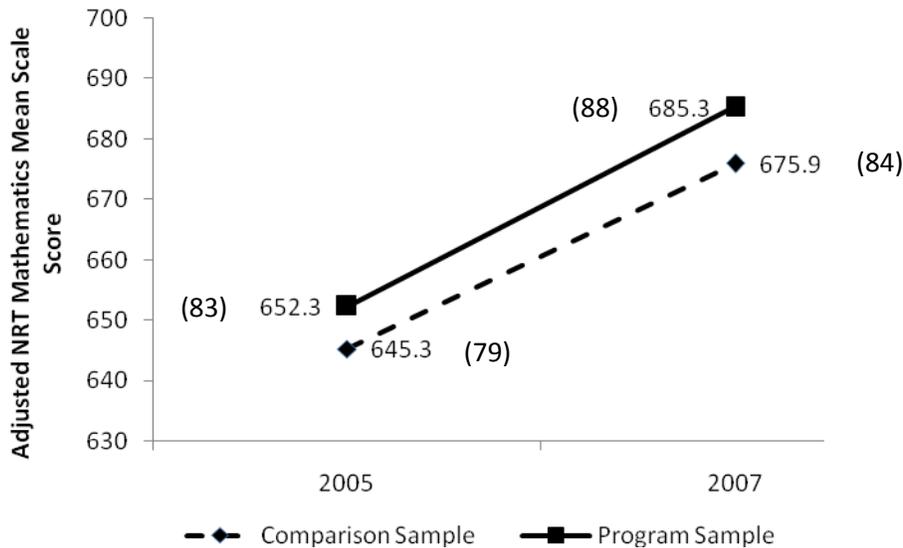


Figure 5

### FCAT NRT Mathematics Achievement Results for Students in the Two Samples

In the analyses presented thus far, the reading and mathematics results for students in the Program Sample were compared to those of their matched counterparts in the Comparison Sample. These analyses did not take into account the difference between the two models implemented within the EFL program. To investigate the potential differential impact of the model of the EFL program, several auxiliary analyses were carried out. The Program Sample was split in two subsamples: Program A Subsample and Program B Subsample. The former contained those students (220 in all) who attended the 18 schools that implemented Model A of the EFL program, and the latter contained 142 students in 10 schools participating in Model B of the program. Because the Comparison Sample was constructed to match students in the Program sample on demographic and previous academic characteristics at student and school levels, no additional matching sam-

ples were created. Instead, the Comparison Sample was split into two subsamples containing students who were originally matched with students in the Program A and B Subsamples. As in the overall analyses of the student academic achievement in reading and mathematics, the three-level HLM was used to compare achievement growth trajectories for students in the Program and Comparison Subsamples for reading and mathematics separately.

### *Auxiliary Analyses Results*

The results of the auxiliary analyses of student academic achievement during the 2005-2007 period indicated that the achievement growth trajectories were similar for students in both subsamples. Generally, students in both Program Subsamples (A and B) had higher beginning (2005) levels of academic achievement and exhibited somewhat higher rates of achievement growth during the 2005-2007 period than did their counterparts in the Comparison Subsamples. The differences in the adjusted mean rates of academic achievement growth were not statistically significant. These results are shown in Table 3 for reading and in Table 4 for mathematics achievement results.

Table 3

### *Reading Achievement Growth Parameters for Students in the two Subsamples*

|         | Adjusted Mean Scale Score (National Percentile) |            |            |            | Adjusted Mean Rate of Annual Growth |            |
|---------|---|------------|------------|------------|-------------------------------------|------------|
|         | 2005  |            | 2007       |            |                                     |            |
|         | Program   | Comparison | Program    | Comparison | Program                             | Comparison |
| Model A | 645.7 (72)                                      | 642.1 (69) | 694.3 (90) | 689.9 (89) | 24.3                                | 23.9       |
| Model B | 640.3 (67)                                      | 638.1 (65) | 690.7 (89) | 685.7 (86) | 25.2                                | 23.8       |

Table 3 shows that that the adjusted mean annual rates of growth in reading achievement were higher for students in both program subsamples compared with those of students in the corresponding comparison subsamples. The differences in the average rates of growth (0.4 scale score points per year for Model A vs. the Comparison and 1.4 scale score points per year for Model B vs. the corresponding Comparison Subsample) were not statistically significant.

Similarly, Table 4 shows that that the adjusted average annual rates of growth in mathematics achievement were higher for students in the program subsamples (both A and B) compared with those of students in the corresponding comparison subsamples. The differences in the adjusted mean rates of growth (1.4 scale score points per year for both Model A and Model B vs. their corresponding Comparison Subsamples) were not statistically significant.

Table 4

*Mathematics Achievement Growth Parameters for Students in the two Subsamples*

|         | Adjusted Mean Scale Score |            |            |            | Adjusted Mean Rate of Annual Growth |            |
|---------|---------------------------|------------|------------|------------|-------------------------------------|------------|
|         | 2005                      |            | 2007       |            |                                     |            |
|         | Program                   | Comparison | Program    | Comparison | Program                             | Comparison |
| Model A | 655.1 (85)                | 648.5 (81) | 687.9 (90) | 678.5 (85) | 16.4                                | 15.0       |
| Model B | 645.4 (79)                | 641.4 (77) | 680.8 (86) | 673.9 (83) | 17.7                                | 16.4       |

*Summary for Evaluation Question 2.* An analysis of the reading and mathematics achievement during the 2005-2007 period revealed that students in the program performed at similar levels of achievement on the reading and mathematics components of the FCAT NRT compared to demographically and academically comparable students. The annual rates of reading and mathematics achievement growth were also similar for the program and comparison students. In addition, the model of the EFL program did not have a differential impact on students' academic achievement growth in both academic disciplines during the 2005-2007 time interval.

*Evaluation Question 3: What are the Teachers' Views on the Implementation of the Program?*

Teachers in the EFL program were surveyed. Of the 236 questionnaires distributed to the EFL teachers in the 33 selected schools, 203 completed questionnaires were return (86% return rate). The Teacher Questionnaire inquired about the implementation of the EFL program in the respondents' schools. It asked about teachers' opinions concerning the effectiveness of EFL program for acquiring or maintaining Spanish language skills. In addition, the questionnaire inquired about teachers' views on possible ways to improve the EFL program.

Teachers were asked to report the length of time of weekly instruction in Spanish received by students who participated in the EFL program. Teachers working in schools implementing Model A of the program in 2006-07 reported an average of 4.1 hours of weekly instruction, while those from Model B schools reported an average of 5.4 hours. The responses from Model A teachers ranged from 1 to 8 hours, whereas those from Model B teachers ranged from 1 to 13 hours.

Teachers in the EFL program were asked whether Spanish instruction was stopped or reduced to prepare students for the 2007 FCAT or the Stanford Achievement Test (administered to students in grades 1 and 2 in the Miami-Dade County Public Schools). Of the 203 respondents, 60 (30%) answered affirmatively. Of those, 23 teachers indicated that the Spanish instruction was stopped for an average of 10 days for these purposes. Another 37 teachers stated that the Spanish instruction was not stopped, but the amount of instructional time in Spanish was reduced to an average of two hours per week for an average of five weeks. When asked whether Spanish instruction was interrupted or reduced for any other reason, only nine of the 203 respondents (4%) answered

affirmatively. Among the reasons mentioned more than once was the need to proctor the FCAT and to administer the Comprehensive English Language Learning Assessment.

Teachers were also asked to appraise how well the EFL program operates in their schools. Of the 200 teachers who answered this question, 66% thought that the program functions above average or very well, 27% thought that the program functions at the average level, and only 7% believed that the program functions poorly or at a below average level in their schools. When the same question was asked in 2004, the proportions of respondents who answered in the same categories were virtually identical.

Teachers provided information concerning a perceived effectiveness of the program in both 2004 and 2007. This information is shown in Table 5.

Table 5

*Perceived Effectiveness of the EFL Program*

| The Program is     | <i>Effectiveness of the Program in</i>                       |              |  |              |
|--------------------|--|--------------|--|--------------|
|                    | Improving Spanish Skills of Native Spanish-speaking Students |              | Helping Students for whom Spanish is not a Native Language to Acquire Spanish Skills |              |
|                    | 2004 (n=243)   | 2007 (n=201) | 2004 (n=225)   | 2007 (n=183) |
| Not Effective      | 1%   | 6%           | 8%   | 11%          |
| Somewhat Effective | 10%  | 6%           | 28%  | 28%          |
| Effective          | 33%  | 37%          | 34%  | 36%          |
| Very Effective     | 56%  | 52%          | 30%  | 26%          |

In addition to rating items, the teacher questionnaire had two open-ended questions. The first question inquired about areas of the EFL program that can be improved, while the second asked about ways to improve them. Altogether, 188 (93%) teachers replied to the first open-ended question. Of those, 58 teachers (31%) expressed the need for better instructional materials (books, CD's, encyclopedias, grammar support books, maps, and consumable workbooks), 43 respondents (23%) stated that the schedule needs to be improved because it does not give EFL teachers enough time for instructional planning, 15 teachers (8%) perceived the need for more oral or writing skills development training, and 13 teachers (7%) thought that specific defined criteria for student entrance to the EFL program need to be established. In addition, ten respondents (5%) indicated that students from non-Spanish language background need additional support, and an equal number of teachers suggested extending the Spanish language instruction within the EFL program to other academic subjects.

Of the 179 teachers who offered their opinions about possible ways to improve the program, 52 (29%) suggested purchasing better instructional materials for students, 25 (14%) advised providing more funds to EFL teachers, 20 (11%) indicated that schools need to help parents to get more involved with the EFL program, 15 teachers (8%) suggested improving school schedules to pro-

vide for more planning/teaching time, and 12 respondents (7%) suggested improving cooperation among teachers, school administrators, and district program staff. In addition, 35 teachers (20%) stated that the EFL program in their schools functioned well and did not need improvements. (Only opinions expressed by at least 10 respondents in their replies to the open-ended questions are presented above.)

## Conclusions and Recommendations

The results of the annual Spanish reading comprehension assessment of students in the EFL program show that students in both models of the program exhibited similar growth in Spanish reading comprehension as measured by Aprenda 2 during the 2004-2007 period. Given that students in Model B of the program had a longer instructional time in Spanish, this result may seem surprising. However, the longer instructional period is used to learn a specific content area (Science or Social Studies) in Spanish. Thus, students in Model B learn content-specific vocabulary in Spanish, but that content may not be assessed by the Aprenda 2 reading comprehension subtest.

The additional instructional time in Spanish proved beneficial for the oral skills development of students from non-Spanish language backgrounds. Such students participating in Model B of the program performed significantly better on the oral component of the Spanish language test than students in Model A, but not as well as students from a Spanish language background participating in either model of the program. Students who had Spanish as a native language exhibited equal levels of Spanish oral skills regardless of the model program in which they participated. Again, this finding is not surprising because such students are likely to have Spanish language exposure at home and in the community.

One of the main objections to bilingual education often stated by its critics is that learning or maintaining the second language will distract students from mastering core academic subjects, such as reading and mathematics. The results of the analysis of the 2005-2007 FCAT NRT results showed no evidence to support this claim. Students in the Program Sample exhibited 2005 through 2007 achievement levels in reading and mathematics that was on par with or higher than that of demographically and academically similar students. In addition, annual learning rates in both academic disciplines were similar for program and comparison students. This finding indicates that learning Spanish and content areas in Spanish did not divert EFL students from learning reading and mathematics. Furthermore, it suggests that the EFL program provided a valuable benefit to participating students: gaining knowledge of a second language while keeping on par with similar students not in the program in core academic subjects. It is reasonable to expect that the knowledge of the second language, if properly maintained throughout the school years, will prove to be a valuable asset to students in their post-graduation endeavors.

The analysis of the teacher survey indicated that, according to teachers' estimates, the length of instructional time in Spanish was shorter than that specified by the program design for Model A and especially for Model B of the program. In addition, almost one-third of the teachers reported interruptions or reductions in instructional time in the program due to preparing students for the FCAT. These findings indicate that the potential positive impact of the program was reduced.

About two-thirds of teachers stated in both 2004 and 2007 that the EFL program operated above average or very well in their schools. In addition, more than one-half of teachers surveyed in both years believed that the program was very effective in improving Spanish skills of students from Spanish language background. On the other hand, in 2004, less than one-third of teachers indicated that the program was very effective in helping students from a non-Spanish language background to acquire Spanish language skills. By 2007, only about one-fourth of teachers thought that the program was very effective in that regard.

Teachers' responses to the survey also indicated that a sizable proportion of respondents perceived the need for better instructional materials and improvement in school schedules to allow for teachers to have more planning time.

In accordance with the results discussed above, the following recommendations are offered:

1. Develop a plan to provide all students who participated in the EFL program throughout elementary school an opportunity to continue learning the second language in middle school.
2. Develop a plan to address language needs of students from non-Spanish language backgrounds within the EFL program with the goal of reducing the gap in Spanish language achievement between them and native Spanish speakers by the end of elementary school.
3. Develop a uniform district policy regarding a temporary halt or reduction in time of the Spanish language instruction within the EFL program to prepare students for the FCAT.
5. Develop a plan to address the needs of teachers in the EFL program for instructional and supplemental materials.

## References

- Harcourt Brace Educational Measurement. (1998). *Aprenda Second Edition Technical Manual*. San Antonio, TX: Harcourt Brace & Company.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, California: Sage Publications, Inc.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of propensity score in observational studies. *Biometrika*, 70 (1), 41-55.
- Sekhon, J. S. (2007). *Multivariate and propensity score matching software with automated balance optimization: the matching package for R*. Retrieved from <http://sekhon.berkeley.edu/matching>

## Appendix A

### List of Participating Schools

|      |                                       |
|------|---------------------------------------|
| 0071 | Thomas, Eugenia B. Elementary School  |
| 0121 | Auburndale Elementary School          |
| 0271 | Bent Tree Elementary School           |
| 0481 | Bright, J. H. Elementary School       |
| 0671 | Calusa Elementary School              |
| 0721 | Carver, G. W. Elementary School       |
| 0831 | Pepper, Claude Elementary School      |
| 1001 | Coral Park Elementary School          |
| 1481 | DuPuis, John G. Elementary School     |
| 1641 | Emerson Elementary School             |
| 1721 | Everglades K-8 Center                 |
| 1761 | Fairchild, David Elementary School    |
| 1801 | Fairlawn Elementary School            |
| 2111 | Hialeah Gardens Elementary School     |
| 2331 | Hadley, Charles R. Elementary School  |
| 2511 | Hurston, Zora Neale Elementary School |
| 2661 | Kensington Park Elementary School     |
| 2781 | Kinloch Park Elementary School        |
| 3111 | Matthews, Wesley Elementary School    |
| 3421 | Milam, M.A. K-8 Center                |
| 3741 | North Beach Elementary School         |
| 3901 | North Hialeah Elementary School       |
| 3981 | North Twin Lakes Elementary School    |
| 4241 | Palm Lakes Elementary School          |
| 4261 | Palm Springs Elementary School        |
| 4281 | Palm Springs North Elementary School  |
| 4761 | Royal Palm Elementary School          |
| 5021 | Sheppard, Ben Elementary School       |
| 5051 | Graham, Ernest R. Elementary School   |
| 5101 | Smith, John I. Elementary School      |
| 5441 | Sylvania Heights Elementary School    |
| 5601 | Twin Lakes Elementary School          |
| 5991 | Wyche, Charles D. Elementary School   |

Appendix B

2004 Teacher Questionnaire

*The Office of Evaluation and Research is conducting an evaluation of the Extended Foreign Language (EFL) program. Your participation in the teacher survey is very important to the evaluation and program improvement efforts. Please answer all the questions below.*

1. School Name \_\_\_\_\_ Work Location \_\_\_\_\_

2. How many years have you been teaching in grades K-12?

First year \_\_\_\_\_ Years (include the current school year)

3. Of those, how many years have you been teaching in the EFL program?

First year \_\_\_\_\_ Years (include the current school year)

4. How far did you get in formally studying the Spanish language (grammar, writing, reading)?

In an English-speaking country

In a Spanish-speaking country

Elementary/Middle school courses

Elementary/Middle school

High school courses

High school

College courses

College courses

None

5. Indicate the highest degree in Spanish (creative writing, literature, etc.) or Spanish teaching you received.

In an English-speaking country

In a Spanish-speaking country

Bachelor's

Bachelor's

Master's

Master's

Specialist/Doctorate

Specialist/Doctorate

None

None

6. What is (are) your area(s) of certification?

\_\_\_\_\_  
\_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_

**PLEASE CONTINUE ON THE OTHER SIDE**

7. What is your current teaching assignment (check all that apply)?

- Classroom teacher       ESOL       Spanish-S       Spanish-SL

8. What subject area(s) do you teach in the EFL program (check all that apply)?

- Language Arts       Science       Social Studies       Mathematics

Other (explain) \_\_\_\_\_

9. How well is the EFL program functioning in your school?

- Poorly       Below Average       Average       Above Average       Very well

10. How effective is the EFL program in improving Spanish skills of native Spanish-speaking students?

- Not very effective       Somewhat effective       Effective       Very effective

11. How effective is the EFL program in helping students for whom Spanish is not a native language to acquire Spanish skills?

- Not very effective       Somewhat effective       Effective       Very effective

12. What areas of the EFL program in your school can be improved?

---

---

---

---

---

---

13. What needs to be done to improve these areas?

---

---

---

---

---

---

**THANK YOU VERY MUCH FOR COMPLETING THIS QUESTIONNAIRE!**  
**Please send it to: 9020, Office of Evaluation and Research, Alex Shneyderman**

Appendix C

2007 Teacher Questionnaire

*The Office of Program Evaluation is conducting an evaluation of the Extended Foreign Language (EFL) program. Your participation is a very important part of the evaluation and will assist in the program improvement efforts. Please answer all the questions below.*

1. School Name \_\_\_\_\_ Work Location \_\_\_\_\_

2. Approximately how many hours per week do students participating in the EFL program in your school receive instruction in Spanish?

Less than 1 hour     1 hour     1.5 hours     Other (specify) \_\_\_\_\_ hours

3. Did the Spanish instruction within the EFL program in your school temporarily stop or was the amount of instructional time reduced to prepare students for the 2007 FCAT?

Yes, the amount of Spanish instruction was stopped for approximately \_\_\_\_\_ days.

The Spanish instruction did not stop, but the instructional time was reduced to approximately \_\_\_\_\_ hours per week for \_\_\_\_\_ weeks.

No, Spanish instruction continued the same.

4. Did the Spanish instruction within the EFL program in your school temporarily stop or was the amount of instructional time reduced for any reason **other than** to prepare students for the 2007 FCAT?

Yes (explain) \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_

The Spanish instruction was stopped for approximately \_\_\_\_\_ days.

The Spanish instruction did not stop, but the instructional time was reduced to approximately \_\_\_\_\_ hours per week for \_\_\_\_\_ weeks.

No, Spanish instruction continued the same.

**PLEASE CONTINUE ON THE OTHER SIDE**

