

Evaluation of One Secondary Computer-Based Program Used with English Language Learners

Dr. Aleksandr Shneyderman
Research Services
Miami-Dade Public Schools

May 24, 2013

Introduction

This evaluation was conducted at the request of the Division of Bilingual Education and World Languages, which requested that the Office of Assessment, Research, and Data Analysis examine the English language acquisition outcomes of students who participated in *Achieve 3000*, a computer based instructional program. *Achieve 3000* is a differentiated online instructional program intended to enable continued student reading growth. The program is designed to adjust reading materials to students' reading levels, providing more challenging content as students' reading levels improve.

Evaluation Design

This section describes sampling procedures used to select schools and students. In addition, it addresses the outcome measures used in the evaluation and describes the data analyses performed.

Student Selection

Because *Achieve 3000* is implemented in virtually all secondary schools and K-8 centers in the District, it was not possible to select a sufficiently large sample of students not participating in the program. Instead, all students in grades 6-12 who participated in the program during 2011-2012 were selected. Then, the length of student participation in the program was used to split the set of students into two approximately equal groups. The first group, for which the length of program participation during the 2011-2012 was at least 13.62 hours, constituted the Program Sample. The Comparison Sample included all students who participated in the program for less than 13.62 hours.

The demographic and academic achievement characteristics of the two samples are shown in Table 1. The table shows that the student groups were reasonably well matched demographically. In addition, Program and Comparison Samples were well matched in terms of their 2011 mean scale scores on the Comprehensive English Learning Assessment (CELLA) for students in grades 9-12. On the other hand, grade 6-8 students in the Program Sample had lower mean scale scores than their counterparts in the Comparison Sample in all three CELLA areas. These differences were taken into account statistically as explained in the *Data Analysis* section.

Table 1

Demographic and Achievement Characteristics of the two Samples

	Grades 6-8		Grades 9-12	
	Program	Comparison	Program	Comparison
<i>Mean time of Program Participation</i>	35.7	7.2	30.6	6.7
<i>Percentage of Students who are</i>				
Hispanic	80.6	86.4	85.4	76.1
Eligible for the FRL program	90.6	89.9	87.3	85.4
<i>Mean Scale Scores on the 2011 CELLA</i>				
Listening/Speaking	692.7	703.4	711.6	710.7
Reading	710.0	715.2	739.8	738.7
Writing	698.1	704.5	712.5	711.0

Outcome Measures

Student results on the 2011 and 2012 CELLA were used to examine the English language acquisition outcomes of students in the Program and Comparison Samples. CELLA is a four-skill language proficiency assessment that tests students' listening, speaking, reading, and writing skills. The results are provided as scale scores in the three domains: oral (listening/speaking), reading and writing. CELLA is administered at four different levels. For the analyses in this evaluation, 2012 CELLA results for levels C (grades 6-8) and D (grades 9-12) were used.

Data Analysis

The General Linear Model (GLM) was used to compare the 2012 CELLA mean scale scores for students in the Program and Comparison Samples. The analyses were carried out separately for each of the grade level groupings of students, and independently for each of the three areas in which CELLA scale scores are reported: Listening/Speaking, Reading, and Writing. The 2011 CELLA scores in each modality were used as covariates while the program participation dichotomous indicator was used as a fixed factor. Prior to carrying out the GLM analysis, an assumption of the equality of slopes of regression lines of the 2012 CELLA outcomes on the 2011 outcomes had to be ascertained. It turned out that this assumption was violated in several cases. In those cases, an adjusted procedure comparing the 2012 CELLA outcomes for the low, medium, and high levels of the 2011 CELLA results was conducted. The low, medium, and high levels were defined as the mean minus one standard deviation, the mean, and the mean plus one standard deviation, respectively. All separate analyses were carried out at the .05 level of

statistical significance. The number of students whose scores were used in separate analyses varied from 2106 to 3349.

Results

The results of the statistical analyses are presented in Table 2. In this table, the 2012 adjusted mean scale scores whose differences were found to be statistically significant are shown in bold.

Table 2
2012 CELLA Adjusted Mean Scale Scores

The 2011 CELLA level	Grades 6-8		Grades 9-12	
	Program	Comparison	Program	Comparison
	<i>Listening/Speaking</i>			
Low	698.5	691.3	712.0	704.6
Medium	730.1	725.0	734.4	729.2
High	761.4	758.4	757.4	754.5
	<i>Reading</i>			
	Program	Comparison	Program	Comparison
	<i>Writing</i>			
Low	714.8	708.9	756.2	751.7
Medium	737.1	734.0		
High	759.3	759.0		
	<i>Writing</i>			
	Program	Comparison	Program	Comparison
	<i>Listening/Speaking</i>			
Low	701.6	696.4	727.5	724.9
Medium	726.3	722.9		
High	750.6	749.0		

Note: Since the homogeneity of slopes assumption was satisfied in the analyses of the 2012 CELLA reading and writing performance for students in grades 9-12, only an overall test was performed in each of these subject areas.

Table 2 shows that the 2012 CELLA adjusted mean scale scores of students in the Program Sample were significantly higher than those of students in the Comparison Sample for all but two comparisons. Specifically, the differences between the Program and Comparison Sample 2012 CELLA reading and writing adjusted mean scale scores of the Grade 6-8 students who scored at the high levels on the 2011 CELLA were not found to be statistically significant.

Because the sample sizes were very large (at least 2106), even small differences are likely to be found statistically significant. Therefore, it is important to examine the differences that were found to be statistically different in terms of a standardized effect size. In these analyses, the statistically significant differences expressed in terms of Cohen's d (the difference between the means expressed in standard deviation units) range from a low of .05 to the high of .14. This entire range of values falls under .2, a value that is generally considered a small effect size.

Discussion

In this evaluation, the 2012 CELLA outcomes were compared for students in the Program and Comparison Samples while adjusting for the students' 2011 CELLA results. It was found that the students in the Program Sample had higher 2012 CELLA adjusted mean scale scores than did students in the Comparison Sample for the majority of comparisons. However, the magnitude of the effect size was small.

It is important to note that students in both student samples participated in the program during the 2011-2012 school year. That is, the comparisons were made between students who differed on the *extent* of their participation in the program. Thus, the findings could be interpreted in the following way: the degree of program participation had a small positive effect on students' English language acquisition results.

Because of the way the student samples were defined, it is possible that the length of student participation in the program was related to students' attitudes toward learning. In that case, students with more positive attitudes toward learning are likely to participate in the program for longer periods. Then, higher English language acquisition results of such students could be explained by their more positive attitudes toward learning and not by their degree of program participation.