

Evaluation of Two Elementary Computer-Based Programs Used with English Language Learners

Dr. Aleksandr Shneyderman
Research Services
Miami-Dade Public Schools

April 18, 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 the two computer-based programs used with English Language Learners (ELLs) in the District. The programs included in the evaluation at the elementary school level were Waterford and Imagine Learning. Waterford is designed as a *reading* program, whereas Imagine Learning as a *language development* program. Waterford is designed to be used with native speakers of English as well as with ELLs and is thought to improve the reading skills of participating students. In contrast, Imagine Learning is designed to provide instruction specifically to ELLs. That instruction is said to improve oral language and overall literacy skills of participating students.

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.

Sampling

In February-March 2013, a survey of school principals was conducted to determine which computer-based programs were used with ELLs during the 2011-2012 school year. The results of the survey were used to select two Program Samples of schools to be included in the evaluation. For each of the two programs, a school participating in the program was selected if (1) no other computer-based program for ELLs was implemented in the school during the 2011-2012 academic year and (2) data on student participation in the program were available. Following these rules, 14 elementary schools were selected as members of the Program Sample for the Waterford program, and 13 other elementary schools were selected as elements of the Program Sample for Imagine Learning.

Following the creation of the two Program Samples, a propensity score analysis was conducted at the school level using the schools in the Program Samples and all other elementary schools that had no computer-based programs used with ELLs during 2011-2012. The following

variables were used in the analysis: the size of the student population, the percentage of ELL students, the percentage of students eligible for the federal free/reduced price lunch program, percentages of students classified as gifted or as Special Education, and percentages of students in the major ethnic/racial categories. Then, for each of the schools in the two Program Samples, an attempt to select a demographically similar school (using the propensity score as a similarity measure) was made. These attempts were successful in selecting 11 schools for the Waterford Comparison Sample and 12 schools for the Imagine Learning Comparison Sample. The lists of program and comparison schools are provided in the Appendix.

For the programs' schools, all participating students who completed at least eight of the program lessons were selected. This criterion was used for both computer-based programs that were included in this evaluation. All these students constituted the student Program Samples, one for each of the two programs. For students in the Program Samples, the mean numbers of hours of program participation were 14.3 for Waterford (ranging from 8 to 97 hours) and 19.2 for Imagine Learning (with the range of between 8 and 78 hours).

For each of the students in the Program Samples, his or her October 2011 grade level and the 2011 outcomes on each of the three domains of the Comprehensive English Language Learning Assessment (CELLA) were identified. In addition, for students who were in grade K during the 2011-2012 school year and who, therefore, did not have 2011 CELLA scores, their February 2012 levels in the English for Speakers of Other Languages (ESOL) program were recorded. Then, for students in grades 1 or above, a multivariate matching algorithm was used to find comparison students from matching schools who would match the program students exactly on their grade levels while minimizing the multivariate distance between program and comparison students on the 2011 CELLA scale scores. For grade K students no 2011 CELLA scores were available; thus, students in the comparison group were matched exactly with the program students on their grade and ESOL levels.

The demographic and academic achievement characteristics of the two samples are shown in Table 1 for each of the two elementary programs. In this table, school-level characteristics reflect the two groups of schools represented in each student sample, and student-level characteristics provide the results of the individual student matches.

The results of the matching process presented in Table 1 show that the student groups were reasonably well matched in terms of the characteristics of the schools they attended. In addition, student groups were reasonably well matched at the individual level except that Program Sample students participating in Waterford had a lower mean 2011 CELLA reading scale score than their counterparts in the Comparison Sample. In addition, Imagine Learning Program Sample students had lower mean scale scores than their counterparts on both 2011

CELLA Listening/Speaking and Reading domains. These differences were taken into account statistically as explained in the *Data Analysis* section.

Table 1

Demographic and Achievement Characteristics of the two Samples by Program

	Waterford		Imagine Learning	
	Program	Comparison	Program	Comparison
School-Level Characteristics	n = 14	n = 11	n = 13	n = 12
Mean Size of Student Population	974	631	748	744
<i>Mean Percentage of Students who are</i>				
Hispanic	84	74	76	64
Eligible for the FRL program	84	76	82	64
ELL	39	29	42	20
Gifted	9	10	11	14
SPED	11	11	8	7
Student-Level Characteristics	n = 201	n = 201	n = 136	n = 136
<i>Percentage of Students who are</i>				
Hispanic	98	97	84	83
Eligible for the FRL program	86	86	83	84
<i>Mean Scale Scores on the 2011 CELLA</i>				
Listening/Speaking	644.5	645.2	643.4	658.5
Reading	513.3	519.4	605.6	611.6
Writing	621.4	621.6	646.6	648.8

Outcome Measures

Student results on the 2011 and 2012 CELLA were used to examine the effects of the programs on students’ English language acquisition. CELLA is a four-skill language proficiency assessment that tests ELL students’ listening, speaking, reading, and writing skills. The results are provided as scale scores in the three domains: oral (listening/speaking), reading and writing.

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 programs and 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 for students in grades 1-5. An

attempt to use students' free/reduced price lunch (FRL), gifted and SPED status as covariates was made. However, these variables were found to not be statistically significant predictors of the 2012 CELLA outcomes once the 2011 CELLA scores were used in the model as covariates; consequently, they were removed from the final model. Students who were in grade K during the 2011-2012 school year did not have 2011 CELLA scores; accordingly, their February 2012 ESOL levels were used as covariates in the GLM procedures. A program dichotomous indicator was used as a fixed factor in the GLM process. All separate analyses were carried out at the .05 level of statistical significance.

Results

The results of the statistical analyses are presented in Tables 2 and 3. In these tables, the numbers in parenthesis indicate the numbers of students whose CELLA results were included in the statistical analyses. The adjusted mean scale scores whose differences were found to be statistically significant are shown in bold.

Waterford Results

The results of the analyses for the Waterford program shown in Table 2 indicate that in most CELLA modalities and grade levels analyzed, the differences between the 2012 CELLA adjusted mean scale scores for Program and Comparison Samples were not statistically significant. An exception was the Listening/Speaking modality. There, the 2012 CELLA adjusted mean scale scores for the Program Sample were significantly higher than those of the Comparison Sample in both grade groupings.

Table 2
Results of the GLM Analysis for Waterford

	2012 CELLA Adjusted Mean Scale Scores	
	Program	Comparison
	<i>Listening/Speaking</i>	
Grade K	647.5 (91)	633.8 (83)
Grades 1-2	682.3 (82)	671.9 (81)
	<i>Reading</i>	
Grade K	643.6 (90)	643.0 (83)
Grades 1-2	646.0 (82)	643.5 (81)
	<i>Writing</i>	
Grade K	669.8 (91)	672.1 (84)
Grades 1-2	670.8 (82)	674.3 (81)

The effect size indices for this comparison, the partial eta squared, were 0.056 for grade K and 0.050 for grades 1-2, which generally are considered small to medium. Partial eta-squared can be interpreted as the percentage of variance in the dependent variable that is attributable to the given effect while holding the covariate constant. Thus, about 5-6% of the variance in the 2012 CELLA Listening/Speaking scores is attributable to the students' program participation when holding the students' 2011 CELLA Listening/Speaking scores for grades 1-2 students or their ESOL levels for grade K students constant.

Imagine Learning Results

The results of statistical analyses for the Imagine Learning program shown in Table 3 indicate that the differences between the adjusted 2012 CELLA scale scores between students in the Program and Comparison Samples were not statistically significant for any of the CELLA scale score modalities and grade level groupings involved in the comparisons.

Table 3
Results of the GLM Analysis for Imagine Learning

	2012 CELLA Adjusted Mean Scale Scores	
	Program	Comparison
	<i>Listening/Speaking</i>	
Grades 1-2	687.3 (76)	687.3 (71)
Grades 3-4	695.4 (58)	700.3 (51)
	<i>Reading</i>	
Grades 1-2	664.8 (76)	663.7 (69)
Grades 3-4	694.2 (58)	696.7 (51)
	<i>Writing</i>	
Grades 1-2	672.6 (76)	671.8 (71)
Grades 3-4	695.9 (58)	701.0 (51)

Discussion

Two different elementary school computer-based programs were included in this evaluation. Program schools included in the evaluation were selected so that any school selected for a Program Sample had only one computer-based program used with ELL students operational during the 2011-2012 school year, so that any potential effects on students' English language acquisition could be attributed to program participation. In addition, results of only those students who completed at least eight hours of the programs' lessons were included in the analyses. This was done in an effort to include in the Program Sample only those students for whom there was a minimum "dosage" of exposure to the program activities.

Schools in the programs were matched with other schools in the District. Additionally, students in selected program schools were matched with students in selected comparison schools. The 2011 and 2012 CELLA results were used to assess the potential program effects. The results of the statistical analyses indicate that program students outperformed comparison students (in the sense of statistically significant differences) for one modality of CELLA, and for one program. Specifically, the 2012 CELLA adjusted Listening/Speaking mean scale scores were significantly higher for students who participated in the Waterford program than those of comparison students. It should be noted that Waterford is designed as a *reading* program, however.

A general finding of this evaluation of two elementary school computer-based programs used with ELLs is the apparent lack of the program effect in the areas assessed, except as noted above. Of course, it is possible that certain programs' potential positive effects were not measured by CELLA, the only instrument used. For example, it is likely that students who participated in the technology programs would be more familiar with and have more positive attitudes toward computer technology. These possible positive effects were not assessed by this evaluation. In addition, it should be noted that the evaluation did not take into account any potential *teacher* effects. So, to the extent that the statistically significant program effects were found, they may be due to the *teacher* and not the *program* effect.

Appendix

Waterford

Program Sample Schools		Comparison Schools	
0201	BANYAN ELEMENTARY	0121	AUBURNDALE ELEMENTARY
0521	BROADMOOR ELEMENTARY	0841	COCONUT GROVE ELEMENTARY
0881	COMSTOCK ELEMENTARY	1241	CUTLER RIDGE ELEMENTARY
1001	CORAL PARK ELEMENTARY	1281	CYPRESS ELEMENTARY
1801	FAIRLAWN ELEMENTARY	1641	EMERSON ELEMENTARY
2351	ENEIDA MASSAS HARTNER ELEM.	2801	LAKE STEVENS ELEMENTARY
2511	ZORA NEALE HURSTON ELEMENTARY	4381	DR HENRY E PERRINE ACADEMY
2661	KENSINGTON PARK ELEMENTARY	5421	SUNSET PARK ELEMENTARY
3261	MIAMI HEIGHTS ELEMENTARY	5521	TROPICAL ELEMENTARY
3421	M.A. MILAM K-8 CENTER	5641	VILLAGE GREEN ELEMENTARY
4241	PALM LAKES ELEMENTARY	5981	DR. EDWARD L. WHIGHAM ELEM.
4721	ROCKWAY ELEMENTARY		
5005	DAVID LAWRENCE JR K-8 CENTER		
5081	SKYWAY ELEMENTARY		

Imagine Learning

Program Sample Schools		Comparison Schools	
0341	ARCH CREEK ELEMENTARY	0041	AIR BASE ELEMENTARY
0661	CARIBBEAN ELEMENTARY	0211	DR. MANUEL C. BARREIRO ELEM.
3181	MELROSE ELEMENTARY	0671	CALUSA ELEMENTARY
3741	NORTH BEACH ELEMENTARY	0681	CAROL CITY ELEMENTARY
3901	NORTH HIALEAH ELEMENTARY	1691	CHRISTINA M. EVE ELEMENTARY
3981	NORTH TWIN LAKES ELEMENTARY	2401	HIBISCUS ELEMENTARY
4401	KELSEY L. PHARR ELEMENTARY	2801	LAKE STEVENS ELEMENTARY
4681	RIVERSIDE ELEMENTARY	3301	MIAMI PARK ELEMENTARY
4741	ROYAL GREEN ELEMENTARY	3381	MIAMI SPRINGS ELEMENTARY
5001	SHENANDOAH ELEMENTARY	5401	SUNSET ELEMENTARY
5041	SILVER BLUFF ELEMENTARY	5641	VILLAGE GREEN ELEMENTARY
5321	SOUTHSIDE ELEMENTARY	5991	CHARLES DAVID WYCHE JR ELEM.
5601	TWIN LAKES ELEMENTARY		