Reading Plus:
An Analysis of Usage and Impact, 2011-12

At A Glance

This analysis of the dose response to Reading Plus examined the reading achievement of students who worked with the application during the 2011-12 school year. The analysis compared participating students' adjusted posttest scores, at each of three levels of usage, to the adjusted posttest scores of a reference group, controlling for initial ability and demographic differences. The findings support those of earlier analyses: The application was found to have a consistent beneficial impact on the achievement of the students who used it. Overall, while the results were found to vary somewhat across grade levels, student achievement appears to have benefitted from the availability of this tool, provided as a supplemental resource to traditional classroom instruction.

Background

Reading Plus is an online application comprised of a series of modules that bolster word recognition and recall; teach phonemic awareness, phonics, and fluency, support reading comprehension and analysis (Taylor Associates, 2012). Modules are also available to provide additional support to struggling readers that strengthens visual motor control and enhances visual attention (Shelly-Tremblay & Eyer, 2009). Secondary students have access to modules that build independent reading skills and confidence to prepare them for high-stakes tests, academic success, and challenges beyond secondary school. Integrated assessment tools adjust material to students' reading levels (Taylor Associates). This purpose of this paper is to analyze the impact of the Reading Plus program on the students within the M-DCPS who used it during the 2011-12 school year, through access provided in class and through the Links to Learning initiative.

Methods

Reading Plus is an application designed to bolster students' reading skills. It may be accessed through the classroom or supplementally through the Links to Learning Initiative. The district's Office of Program Evaluation conducted a study to examine students' usage of Reading Plus and conducted an analysis to gauge its impact on students' achievement scores. The study was guided by a series of questions:

1. **To what extent was Reading Plus used by students during the 2011-12 school year?**
2. **Did students who used the software more frequently score higher on standardized achievement tests than students who were typical users?**

Data were gathered from two sources to address the research questions. The first source of data was usage information provided by the software vendor. The second source of data was student demographic data and assessment data maintained on the district's data warehouse.

- **Usage**

  The sample for the study included all students in grades 3 through 10 who used Reading Plus software during the 2011-12 school year. Students who were not active at the end of the school year were excluded from the analysis. The identifying information in the vendor provided files was first validated against district records, and all records with zero usage were removed. Non-zero usage was sorted within grade and classified in four bands, based on percentile: 1 (0 to 39.99), 2 (40.00 - 59.99), 3 (60.00 - 89.99), and 4 (90.00 - 100.00). These bands were defined to provide for inferential comparisons between targeted percentiles of usage located at the midpoint of each band within the distribution, i.e., Band 1 (20th), Band 2 (50th), Band 3 (75th), and Band 4 (95th). Analyses conducted for this section were limited to descriptive statistics.

- **Impact**

  A predictive correlational design (Tuckman, 1999) was used to gauge the impact of the Reading Plus program on students' achievement. The sample was the same as was used in the analysis of usage except that students who did not have valid pre- and post- test scores at consecutive grades, were excluded from the analysis.

  The results of three different achievement measures were used in this analysis: (a) the Stanford Achievement Test, Tenth Edition (SAT-10); (b) the Florida Comprehensive Assessment Test (FCAT); and, (c) the Florida Comprehensive Assessment Test 2.0 (FCAT 2.0). The measures were administered as indicated in Table 1, which separately lists for each posttest (2012) grade level, the pre and post test for reading.

  The SAT-10 is a standardized norm-referenced test designed to measure students' performance in comparison to a national normative sample. Students' performance is measured in scale scores that are equal units of achievement that vertically align across grades, are amenable to mathematical manipulation, and specifically designed to compare individuals and groups. The SAT-10 is administered locally to all students in Grades K, 1, and 2 during the spring of each school year.

Table 1. Reading Achievement Test Administration by Grade

Grade (2012)	Pretest (2011)	Posttest (2012)
2	SAT-10	SAT-10
3	SAT-10	FCAT 2.0
4	FCAT2.0	FCAT 2.0
5	FCAT2.0	FCAT 2.0
6	FCAT2.0	FCAT 2.0
7	FCAT2.0	FCAT 2.0
8	FCAT2.0	FCAT 2.0
9	FCAT2.0	FCAT 2.0
10	FCAT2.0	FCAT 2.0

Note. The achievement tests listed above were administered to students during spring of the years indicated.

The FCAT 2.0, a criterion referenced test designed to measure students' mastery of the state's new Next Generation Sunshine State Standards (NGSSS), is presently the primary accountability measure used by the state of Florida. It was administered statewide to students in Reading (Grades 3 through 10) during April of each school year. Students' performance on FCAT 2.0 is measured in scale scores (i.e., equal units of achievement amenable to mathematical manipulation and specifically designed to compare individuals and groups) and reported in achievement levels that range from 1 (low) to 5 (high).

The analysis compared students' adjusted posttest scores at each of the three levels of usage to their adjusted posttest score at the level of the reference group, controlling for their initial ability and demographic differences. The design was applied in two stages. First, separate stepwise regression analyses, conducted at each grade, were used to predict the influence of demographic characteristics (i.e., gender, ethnicity, free/reduced lunch eligibility, English language learner status, special education status, and age relative to grade) on the students' pre- and post-test scores. The predicted scores produced by this process were then used to create residual scores for each student, which were added to the mean predicted scores at each grade, resulting in rescaled "adjusted" pre- and post- test scores, which accounted for demographic differences.

Second, separate full regression analyses, conducted at each grade, were used to compare the difference in the adjusted posttest scores for students at each of three usage levels (i.e., bands, 1,3, and 4) with the adjusted posttest scores for students at the typical usage level (i.e., band 2), controlling for the influence of the adjusted pretest. Interactions between each of usage levels and the adjusted pretest were also defined to account for the possibility that the effect of usage varied with the level of the adjusted pretest.

Results

- **Usage**

Non-zero usage was sorted within grade and classified in four bands, based on percentile, with midpoints as follows: 1 (20th), 2 (50th), 3 (75th), and 4 (95th). These bands were centered at the 20th, 50th, 75th, and 95th percentiles, respectively. Tables 2 lists for each grade the total number of students and the hours used by at student at the midpoints of each of the four bands of usage.

The table shows that Reading Plus was used by over 15,000 students per grade at most grade levels during the 2011-12 school year. However, half of the students used the software for less than 15 hours all year (or about 1½ hours per month), and 95% used it for fewer than 55 hours (or about 1½ hour per week).

Table 2. Total Hours of Usage

Grade	N	Percentiles			
		20	50	75	95
03	18,519	4.00	15.40	29.20	55.90
04	16,806	3.70	13.50	26.40	52.20
05	17,140	3.60	14.90	27.40	49.30
06	19,642	3.50	12.60	24.90	49.10
07	17,798	2.80	10.00	20.80	41.70
08	17,751	3.30	10.70	20.80	41.70
09	15,650	3.30	12.00	23.40	44.80
10	13,624	3.70	12.40	23.50	45.30
ALL	136,930	3.48	12.70	24.60	47.63

- **Impact**

The predictive correlational design was applied in a two stage process. First, least squares regression was used to adjust students' pre- and post- test scores to account for the influence of demographic characteristics. Second, the students' adjusted post- test scores were compared at different levels of usage, controlling for baseline achievement.

  - **Demographic Adjustment**

    Separate least squares regression analyses with stepwise selection conducted at each grade were used to examine the influence of demographic characteristics (i.e., gender, ethnicity, free/reduced lunch eligibility, English language learner

status, special education status, and age relative to grade) on students' pre- and post- test scores. Table 3 presents the results of these analyses.

Table 3 shows that generally students who are classified as Black, Disabled, or English Language Learners tend to score lower than students not so classified, while students who are Gifted or who are in/not eligible for Free/Reduced Price Lunch tend to score higher than students not classified as such. Additionally, students who are older also tend to score lower than those who are younger (except at the lowest grades when age is a developmental advantage). These relationships tend to be fairly consistent with certain exceptions (e.g., Female) across the grades, regardless of whether the pre- or the post- test is being predicted.

Table 3. Regression Coefficients used for Demographic Adjustment

Predictor	Grade							
	3	4	5	6	7	8	9	10
Pretest								
Intercept	617.71	323.15	329.07	309.91	318.54	323.43	311.78	316.17
Black	-19.60	-20.53	-23.00	-21.85	-26.39	-19.94	-15.12	-16.69
Disabled	-31.44	-36.73	-39.46	-38.41	-43.83	-38.28	-31.41	-29.04
English Language Learner	-42.35	-50.05	-58.34	-73.49	-76.20	-80.87	-73.62	-61.94
Female	4.16	4.37	2.04	2.31	--	--	3.34	--
Gifted	29.07	46.46	46.66	50.98	54.72	52.24	44.24	42.60
Over Age	1.75	-6.08	-7.39	-8.86	-13.09	-12.68	-10.76	-11.14
Free/Reduced Price Lunch [a]	10.70	16.93	14.45	19.57	18.58	19.94	11.89	11.99
Posttest								
Intercept	205.05	215.73	223.82	224.84	232.08	238.28	240.56	245.81
Black	-10.93	-9.08	-9.82	-9.41	-8.05	-7.86	-7.30	-7.62
Disabled	-16.41	-13.98	-15.68	-13.01	-13.12	-13.27	-12.23	-10.49
English language learner	-17.62	-13.88	-16.41	-17.98	-21.72	-23.77	-23.29	-18.85
Female	0.82	1.21	1.64	1.95	--	1.60	--	-2.53
Gifted	16.58	16.11	17.96	19.28	19.79	20.28	18.68	19.08
Over Age	--	-3.03	-3.43	-4.50	-4.95	-5.70	-6.06	-7.03
Free/Reduced Price Lunch [a]	7.05	6.47	5.67	7.60	7.35	7.77	5.37	5.44

Note. All coefficients shown are statistically significant ($p < .01$). Cells displayed as dashes represent coefficients that are not statistically significant ($p < .05$).
[a] Reverse coded.

- **Dose Response**

Therefore, the analysis compared students' adjusted posttest scores at each of the three levels of usage to their adjusted posttest score at the level of the reference group, controlling for their initial ability and demographic differences. Separate least squares regression analyses without selection conducted at each

grade were used to examine the influence of each of the of the usage bands and the adjusted pretest on the adjusted posttest. Three dummy variables were created for the first, third, and fourth usage bands with the second usage band served as the reference group. Table 5 lists for each predictor the statistics for the unstandardized coefficients, the standardized coefficients, and the results of tests of their statistical significance.

Table 5. Regression Analysis of Student Usage

Grade	Predictor	Unstandardized Coefficients M	SE	Standardized Coefficients	t
3	Intercept	203.50	0.24		852.62***
	Band 1	-0.80	0.29	-0.02	-2.71**
	Band 3	1.12	0.31	0.03	3.66***
	Band 4	2.65	0.41	0.05	6.55***
	Adjusted Pretest	0.32	0.00	0.59	93.18***
4	Intercept	214.77	0.21		1018.76***
	Band 1	-0.68	0.26	-0.02	-2.62**
	Band 3	1.47	0.27	0.04	5.41***
	Band 4	3.05	0.36	0.06	8.39***
	Adjusted Pretest	0.23	0.00	0.64	104.97***
5	Intercept	223.37	0.21		1041.61***
	Band 1	-0.75	0.26	-0.02	-2.84**
	Band 3	1.23	0.28	0.03	4.46***
	Band 4	1.90	0.37	0.03	5.14***
	Adjusted Pretest	0.25	0.00	0.67	116.09***
6	Intercept	224.06	0.19		1165.34***
	Band 1	-1.08	0.24	-0.03	-4.55***
	Band 3	2.10	0.25	0.06	8.48***
	Band 4	3.07	0.33	0.06	9.29***
	Adjusted Pretest	0.25	0.00	0.69	131.29***
7	Intercept	230.37	0.20		1128.36***
	Band 1	-1.13	0.25	-0.03	-4.51***
	Band 3	1.32	0.26	0.04	5.02***
	Band 4	2.04	0.35	0.04	5.84***
	Adjusted Pretest	0.25	0.00	0.69	124.01***
8	Intercept	237.62	0.21		1150.02***
	Band 1	-1.13	0.26	-0.03	-4.42***
	Band 3	1.55	0.27	0.04	5.84***
	Band 4	2.21	0.35	0.04	6.24***
	Adjusted Pretest	0.27	0.00	0.70	128.47***

(*table continues*)

Grade	Predictor	Unstandardized Coefficients		Standardized Coefficients	t
		M	SE		
9	Intercept	238.44	0.23		1045.09***
	Band 1	-1.02	0.28	-0.03	-3.61***
	Band 3	0.95	0.29	0.03	3.24**
	Band 4	2.08	0.39	0.04	5.28***
	Adjusted Pretest	0.27	0.00	0.67	105.30***
10	Intercept	243.38	0.24		1008.74***
	Band 1	-0.90	0.30	-0.03	-3.02**
	Band 3	0.66	0.31	0.02	2.12**
	Band 4	2.12	0.42	0.04	5.08***
	Adjusted Pretest	0.29	0.00	0.66	98.07***

Note. The criterion variable is the Adjusted Posttest. Dichotomous predictors are defined for three usage bands based on percentile of usage within each grade: 1 (20th), 3 (75th), and 4 (95th). The reference group is 2: (50th). Adjusted Pretest is grand mean centered within grade and takes into account student demographic characteristics.
** $p < .01.$ *** $p < .001.$

The table shows positive statistically significant coefficients for Band 3 and Band 4 and statistically significant negative coefficients for Band 1 at each grade. For example, in Grade 6, students at the 75th percentile of usage (Band 3) scored 2.10 scaled score points higher than typical users, while students at the 95th percentile of usage (Band 4) scored 3.07 scaled score points higher than typical users, once their initial ability and demographic characteristics were taken into account.

This pattern indicates that increased usage of the software is associated with higher test scores once students' baseline achievement and demographic characteristics are taken into account. No meaningful interactions between the adjusted pretest and the usage bands were found. The strength of the differences at each band within each grade (not shown) was weak according to Cohen's (1988) classification for the effect size of adding a term to a regression analysis: .14 (weak), .36 (moderate), .58 (strong).

Discussion

The results for the 2011-12 school year on the Reading Plus application support those found in earlier analyses. The application was found to have a consistent beneficial impact on the achievement of the students who used it. Overall, while the results were found to vary somewhat across grade levels, student achievement appears to have benefitted from the availability of this tool, provided as a supplemental resource to traditional classroom instruction.

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shelley-Tremblay, J., & Eyer, J. (2009). Effect of the Reading Plus program on the reading skills of second graders. *Journal of Behavioral Optometry, 20* (3), 1-66.

Taylor Associates (2012). *Reading Plus*. Retrieved November 27, 2012 from http://www.readingplus.com/

Tuckman, B.W. (1999). Conducting educational research. Belmont, CA: Wadsworth Group/Thompson Learning.