Miami-Dade County Public Schools

Third Grade Summer Reading Camps Evaluation

2009-10

November 2009

Principal Evaluator/Author:
Steven M. Urdegar, M.B.A., Ph.D.
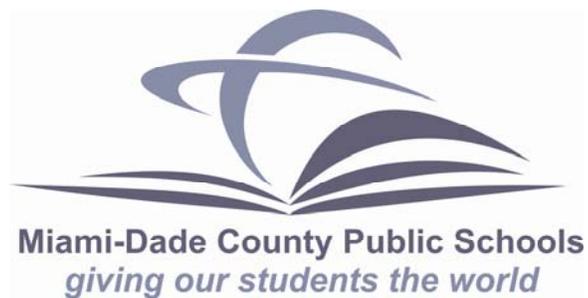
TABLE OF CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

The fundamental goal of the Reading Camp program is to bolster the students' reading skills and thus prepare them for the next opportunity to pass either the FCAT-SSS or an alternative assessment that will enable third grade students who are scheduled for retention to be promoted to the fourth grade. The Office of Program Evaluation undertook an evaluation to gauge whether students who completed the program exhibited improved performance on the alternative assessment when compared to a comparison group of students who did not participate. An examination of students' spring to summer growth on the alternative assessments and the extent to which that growth was influenced by participants' demographic characteristics was also conducted. Alternative assessments were offered to students before and after the Reading Camps. The Iowa Test of Basic Skills, Edition C (ITBS-C), administered at the end of the regular school year, served as the pretest. The Stanford Achievement Test, Tenth Edition (SAT-10), administered at the end of summer school, served as the posttest.

Students who completed the program were found to score 9.5 points higher on the posttest than students who did not participate in the program, when their pretest scores and demographic characteristics were taken into account. Students who completed the program were also found to be 3.9 times more likely to have met the cutoff for promotion on the posttest than students who did not participate in the program, when their pretest scores and demographic characteristics were taken into account. The results were obtained by using a quasi-experimental design to adjust students' pre- and posttest scores to account for the effects of demographic characteristics, then comparing the groups' adjusted posttest scores, while controlling for students' adjusted pretest scores. A regression analysis compared the groups' adjusted posttest scores and estimated the impact of the Reading Camps on the students' adjusted posttest scores, while controlling for their adjusted pretest scores.

Significant achievement status gaps were found for students who were male, black, FRL eligible, ELL, and placed in SPED that mirrored the results of other studies (e.g., Eamon, 2005; Levitt et al. 2005; McLoyd, 1990). Though strong achievement growth was seen for participating students, as a whole, the achievement growth of students designated as ELL and SPED, was significantly lower than that of students who were not so designated. The results were obtained by using a multiple time series design, applied via a repeated measures analysis of variance, to gauge the influence of demographic characteristic on the participants' scores. A mathematical procedure was used to apply a difficulty correction to the pretest and transform its scale to make it comparable to the posttest. A repeated-measures analysis was used to apply the design.

The achievement status gaps and attenuated achievement growth of students who were designated as ELL and SPED, relative to students who were not so designated, suggests the need for additional strategies to benefit those students. The small percentage of non-participants with posttest data limits the generalizability of this study, as the performance of the few students who were included may not be representative.

<div style="background-color:#d9e2bc; text-align:center;">

INTRODUCTION

</div>

The Third Grade Summer Reading Camps were first implemented in the Miami-Dade County Public Schools during the 2003 summer school term as an intensive reading remediation program. The students targeted for enrollment in the camps are primarily third graders who failed to score above Level 1 on the reading subtest of the Florida Comprehensive Assessment Test (FCAT), Sunshine State Standards component and as such are typically barred from advancing to grade 4. The following report details the evaluation of the seventh implementation of the camps (summer 2009) and focuses on the impact of the camps on the students' subsequent performance on the alternative assessment test.

Background

Florida Statutes currently prohibit social promotion and also mandate that students who exhibit academic deficiencies be provided with intensive remediation where particular emphasis is given to the reading proficiency of students in the grades K through 3. Beginning in the 2002-03 school year, districts were required to develop an academic improvement plan for each student who exhibited a deficiency in reading that would "identify the student's specific areas of deficiency in phonemic awareness, phonics, fluency, comprehension, and vocabulary; the desired levels of performance in these areas; and the instructional and support services to be provided to meet the desired levels of performance" (*Public School Student Progression*, 2008).

In the primary grades, the fundamental goal of such a plan is to remediate the student's reading deficiency by the end of grade 3 as evidenced by the attainment of a score of level 2 and above on the reading subtest of the Florida Comprehensive Assessment Test, Sunshine State Standards component (FCAT-SSS) administered in grade 3. The FCAT-SSS is the criterion-referenced portion of the FCAT test battery. The Florida Department of Education (2005) cautions that scoring at level 1 on the reading subtest of the FCAT-SSS may lead to retention in grade 3, unless the student meets one of the following six types of exemptions:

➢ an acceptable level of performance on an alternative assessment;
➢ demonstration, through a teacher-developed portfolio, of the ability to read on grade level;
➢ status as an English Language Learner, with less than two years in English for Speakers of Other Languages programs;
➢ eligibility for special education (SPED) with an individual education plan (IEP) that indicates that it is not appropriate to take the FCAT;
➢ eligibility for SPED without FCAT exemption with an IEP or 504 plan demonstrating receipt of two years of intensive remediation in reading, a continuing deficiency in reading, and a previous retention in grades K through 3; and,
➢ receipt of two years of intensive remediation in reading, a continuing deficiency in reading, and two previous retentions in grades K through 3

The fact that around one-fifth of the state's third graders have historically scored at level 1 during that first year has prompted the state to call for school districts to provide "intensive reading camps" for these students during the summer that serve only to prepare the students for their next opportunity to pass the test (Florida Department of Education, 2005).

The Miami-Dade County Public Schools (M-DCPS) responded to the state's initial suggestion by developing the Third Grade Summer Reading Camps (Reading Camps) to provide students with intensive reading remediation. To ensure uniformity of implementation, the Reading Camp teachers were provided with both in-service training and a detailed curriculum guide. Since its inception, the program has undergone numerous changes in format and length of delivery from a four component model consisting of (a) teacher-directed grade level instruction, (b) Miami Max, (c) Voyager Passport, and (d) Quick Reads to a two component model consisting of (a) Voyager Time Warp Plus and (b) SAT-10 Prep. Despite these variations, evaluations of the program across the years have shown it to have a positive impact on student performance and to increase students' passing rate on alternative assessments.

Description of the Program

The reading camp curriculum that operated this past summer consisted of three components: (a) Voyager Time Warp Plus, (b) Ticket to Read, and (c) SAT-10 Prep. Voyager Time Warp Plus, the first component of the new curriculum is a carryover from the program that operated for the last two years. It is a reading program developed by Voyager Expanded Learning of Dallas, Texas. The program was designed as a summer reading intervention for low-performing students in grade 3. It calls for four hours of reading-related activities per day for four weeks. The activities adhere to a common theme of the Roman Empire. Various materials are used to develop the students' reading fluency through the study of Roman society and history.

> Various print resources are used to develop and enhance reading fluency through the study of the structure of Roman society, engineering feats, and artistic achievements of this great civilization. Skills such as phonological awareness, vocabulary development, and comprehension are taught, modeled, and assessed as students discover links between the past and the present. Teachers will be completely prepared to lead an exciting reading adventure. Easy-to-use curriculum guides, student materials, a theme-based library, hands-on learning activities, assessment materials, even paper and pens ensure that teachers can minimize preparation and maximize instruction. (Department of Language Arts/Reading, personal communication)

The reading-related skills and strategies that are addressed conform to state requirements (Florida Department of Education, 2005). They include (a) fluency, (b) phonological awareness, (c) alphabetic principle, (d) vocabulary development and concept awareness, (e) comprehension and literary response, (f) writing, and (g) listening and speaking (Gomez, 2008).

Ticket to Read is a web-based computer program, which was developed by Voyager Expanded Learning of Dallas, Texas. This is the same company that developed Voyager Time Warp Plus. Ticket to Read is designed to allow students in grades K to 5 to work independently on improving their reading skills. The web-based aspect of the program allows the students to access it at any time from school or home. Students also have their own personal online clubhouse, decorated with items purchased by meeting reading goals and correctly answering comprehension questions. Oral fluency assessment scores are used to place students at one of 16 levels that can accommodate the needs of various readers from those who are struggling to those who are accomplished. Students work independently on fluency, vocabulary, and comprehension Each skill level consists of 30 to 40 reading passages that are designed to appeal to the students. They select the passages that interest them and proceed across skill levels at their own pace. The

program encourages the students through continuous feedback and by providing them with the opportunity to earn virtual prizes for their work (Department of Language Arts/Reading, personal communication).

SAT-10 Prep, is a curriculum that is designed to instruct students in the skill areas assessed by the SAT-10 test, which is the final alternative assessment used at the end of summer to gauge students' eligibility for promotion to grade 4. The district has for the second year scheduled this alternative assessment test during the final days of the summer school term in order to identify the students who merited promotion prior to the start of the school year. The curriculum currently consists of 24 half-days of instruction.

The district's *2009 Summer Services Guide, 2009, Summer Implementation Document* (Department of Summer Services, 2009) defines it simply as "students who were retained in grade 3" (p. 10). Consequently, the students targeted by the Reading Camp program remained primarily the third graders who scored at level 1 on the reading subtest of the FCAT-SSS administered in March of 2009 and did not meet one of the six exemption provisions outlined earlier in this report.

Table 1
Features of the Third Grade Summer Reading Camps

Program Features					
	2005	2006	2007	2008	2009
Hours per day [a]	6 ½	6 ½	6 ½	6	3
Days per week	5	5	5	5	5
Daily schedule	8:45-3:15	8:45-3:15	8:30-3:00[b]	8:30-2:30	8:30-11:30 [b]
Weeks per term	4	4	4	4	5
Hours of instruction per term	120	120	117[b]	110	60
Teacher-student ratio	1:15	1:15	1:15	1:18	1:18
Number of camps	46	45	42	42	28
Student enrollment	5,362	6,167	5,793	2,855	2,171

Note. Each column represents the schedule in effect for that year.
[a] The hours per day includes ½ hour for lunch. [b] In the 2007 program, the Wednesday schedule was 8:30-2:00.

The table provides information on various operational features of the Reading Camp program in each summer that it was implemented during the last five years. Each column represents the schedule in effect for a given year. A review of the table reveals that the greatest number of changes occurred during the last two years depicted in the table. Both of these changes were due primarily to the budget constraints.

The transition from the 2007 program to the 2008 yielded five changes. Enrollment decreased by half, but the length of the day remained roughly the same. The total hours of instruction for the term decreased from 117 to 110 with the teacher-student ratio rising to 1:18. The transition from 2008 to 2009 yielded six changes. The most notable was a decrease by nearly half in the total hours of instruction. This was mainly due to a reduction in the length of the instructional day to three hours including lunch. Enrollment also decreased by one-fourth. Finally, there was a decrease by one-third in the number of sites. A list of the Reading Camps appears in the Appendix Table A.

METHODOLOGY

The Reading Camps are designed to provide intensive reading remediation to third graders, slated for retention due to their performance on the FCAT, and to prepare them to score high enough on an alternative standardized reading assessment to be promoted.

Research Questions

To ascertain the impact of the camps, an evaluation was undertaken by the district's Office of Program Evaluation. The evaluation was guided by a series of questions:

1. **Did students who participated in the Reading Camps score higher on the summer alternative assessment than students who did not participate, when their spring alternative assessment scores and demographic factors were taken into account?**
2. **Were students who participated in the Reading Camps more likely to score high enough on the summer alternative assessment to be promoted than students who did not participate, when their spring alternative assessment scores and demographic factors were taken into account**
3. **Did students who attended the Reading Camps show growth on the alternative assessment from spring to summer; was that growth influenced by their demographic characteristics?**

The first and second research questions were designed to gauge the program's impact. Question 1 examined whether students who completed the program exhibited improved performance on the alternative assessment, when compared to a comparison group of students who did not participate in the program. Question 2 explored whether students who completed the program would be more likely to score high enough on the alternative assessment to be promoted, when compared to students who did not participate in the program. The third research question addressed student growth, and explored the equitable distribution of growth among the various levels of selected demographic subgroups.

Data Sources

The initial source of data for the evaluation of the Reading Camps consisted of a review of documents obtained from the district's Division of Language Arts/Reading, the district's School Choice and Parental Options, and the Florida Department of Education. The latter included the applicable statutes of the Florida School Code. Data from these documents were used in preparing the description of the Reading Camps, which appears in the Description of the Program. To address the aforementioned evaluation question, however, it was necessary to gauge the third graders' performance on the alternative assessment test following their completion of the Reading Camp program.

As previously mentioned, third graders scheduled for retention based on their performance on the FCAT-SSS, or other factors, were provided with multiple alternative assessments that if passed would have allowed them to progress to the fourth grade. The last of these opportunities took the form of a norm referenced instrument administered before the end of the spring semester. This test provided a baseline measure of student performance for students who went on to attend the

Reading Camps, as well as those who did not, and made it possible to address the research questions that are discussed above. The remainder of the methodology will be organized according to those questions.

Program Impact

Design

A quasi-experimental design (Campbell & Stanley, 1963), which involves using pretest and posttest scores to compare the performance of a group that is exposed to an experimental treatment (i.e., program participants), with that of a group that is not (i.e., non-participants), was used to gauge the impact of the camps on students' test scores. The two groups are considered "nonequivalent," because the subjects were not randomly assigned to them, as would have been the case in a true experimental design. Consequently, the pretest scores are used as evidence of the comparability of the two groups prior to the treatment.

Samples

The population for this evaluation consisted of third grade students who were scheduled for retention after having exhausted all alternative assessments for promotion available to them prior to the end of the regular school year. Of the 26,124 students who participated in the statewide assessment and were active in the district at the end of the 2008-09 school year, 21.6% (*n* = 5,638) scored within achievement level 1. These students faced mandatory retention under state statute unless they qualified for one or more good cause exemptions outlined by the state. Table 2 lists the number and percent of students that utilized those exemptions.

Table 2
Good Cause Exemptions to Mandatory Retention
Granted to M-DCPS Third Graders who Scored at Level 1 on the FCAT Reading Subtest, 2008-09

	N	%
English Language Learner [a]	679	23.0
Special Education [b]	838	28.3
Portfolio Assessment	771	26.1
Alternative Assessment [c]	282	9.5
Other [d]	386	13.1
Total	2,956	100.0

Note. Numbers are informational and may contain data entry and categorization errors.
[a] Students with less than two years in an English for Speakers of Other Languages Program [b] Students who were retained once and had two years of remediation [c] ITBS-C offered at the end of the school year [d] Includes students who were retained twice.

Students listed as receiving exemptions based on the Portfolio and Alternative assessments earned passing scores on those tests. The remaining 2,682 students who were still slated for retention were given the opportunity to attend the Reading Camps. The students who elected to attend and went on to complete that program were included in the treatment group. A control was also defined which consisted of the students slated for retention who did not attend the

Reading Camps. Students who participated in the program in earlier years were excluded from both groups, as their involvement would have precluded the isolation of the current program's impact. Table 3 partitions the students slated for retention according to their current and prior participation in the Reading Camps.

Table 3
Crosstabulation of Previous vs. Current Program Exposure Status:
Students Slated for Retention

Previous Exposure	Current Exposure			Total
	Complete	Partial	None	
No	1,872	111	548	2,531
Yes	96	10	45	151
Total	1,968	121	593	2,682

Table 3 shows that of the 2,682 students slated for retention, a total of 2,531 had no previous exposure to the Reading Camps. The treatment pool was comprised of the 1,872 of those students who completed the program. The control pool was comprised of the 548 students who had no current or previous exposure to the program. The other groups were excluded. Of the 1,872 students in the treatment pool 97.5% (n = 1, 825) had valid posttest (i.e., summer alternative assessment) scores. These students constituted the treatment group. Of the 548 students in the control pool, only 16.4% (n = 90) had valid posttest scores. These students constituted the control group. The characteristics of the final sample are described in Table 4. It lists for each subgroup, the percentage of students in the treatment and control group.

Table 4
The Groups' Demographic Characteristics as a Percentage of the Sample

Group	Treatment (n = 1,825)	Control (n = 90)
Sex		
Male	59.8	60.0
Female	40.2	40.0
Race		
Black	43.2	30.0
Hispanic	53.7	60.0
White/Other	3.1	10.0
Free/Reduced Price Lunch		
Eligible	89.2	61.1
non-Eligible	10.8	38.9
English Language Learner		
Current	22.0	12.2
Former	28.7	34.4
Never	49.3	53.4
Special Education		
Disabled	19.2	12.2
non-Disabled	80.8	87.8
Age	9.5	9.5

Note. Numbers are percents except age, which is expressed in years.

Table 4 shows the groups to be similar in terms of age, sex, and special education status. However, the treatment group is more disadvantaged than the control group. Compared to the students in the control group, higher percentages of students in the treatment group are eligible for reduced price lunch, English language learners, and are Black.

Instrumentation

The posttest used for this analysis was the reading comprehension subtest of the Stanford Achievement Test, Tenth Edition, SAT-10, a standardized norm-referenced test designed to measure students' performance in comparison to a national normative sample, and to facilitate comparisons among individuals and groups. It is currently administered by the district to third grade students as an alternative assessment at the end of summer school.

The reading comprehension subtest of the Iowa Test of Basic Skills, Edition C (ITBS-C) was used as the pretest. The ITBS-C is a standardized norm-referenced test designed to measure students' performance in comparison to a national normative sample and to facilitate comparisons among individuals and groups. It is currently administered by the district to third grade students as an alternative assessment at the end of the school year.

Data Analyses

The application of the quasi-experimental designs involved a two phase process that used separate data analyses to adjust students' pre- and posttest scores to account for the effects of demographic characteristics and then compared the groups' adjusted posttest scores, while controlling for students' initial ability as measured by their adjusted pretest scores.

The first phase used regression analysis[1] to separately estimate the effect of selected demographic characteristics (i.e., gender, ethnicity, reduced price lunch eligibility, ELL status, SPED classification, and age relative to grade) on each student's pretest and posttest scale score in order to account for the influence of those factors when contrasting the groups' performance. Regressions were fitted using a full model and forward selection. Predictors that were not significant (p < .05) in either model were removed. A unique estimate was then generated for each student, which was subtracted from the student's score, and the average estimate for all students was added in its place. This process resulted in adjusted pretest and posttest scores for each student that represented the pretest and posttest scores that would have been obtained had the student had the same demographic characteristics as the overall sample.

Status

The question of whether participants scored higher on the alternative assessment than non-participants did was answered by conducting the second phase of the analysis of student

---

[1] Regression analysis is a statistical technique that represents an outcome variable (e.g., posttest) as a sum of a series of explanatory variables (e.g., demographics) in order to "predict" the value of the outcome variable under a variety of conditions. The weights produced by a regression analysis are those that maximize the variation in the outcome variable explained by the predictors. The analysis estimates the degree of influence (weight) of each explanatory variable (predictor) on the outcome. Categorical predictors are represented by a series of dichotomously coded numbers (0 or 1).

outcomes expressed in original scale score units. Regression analysis was used to apply the quasi-experimental non-equivalent control design and to compare the groups' adjusted-posttest scores. It estimated the impact of the program on the students' adjusted posttest scale scores while controlling for students' adjusted pretest scores. Participation in the program (1=yes, 0=no) and adjusted pretest were the main predictors in the model. An interaction term, program x adjusted pretest was also included to account for the possibility that the impact of the program was influenced by the students' adjusted pretest scores.

Promotion

The question of whether participants were more likely than non-participants were to have scored high enough on the alternative assessment to be promoted was answered by reconducting the second phase of the analysis of student performance with the outcome expressed as the dichotomous result of a logical comparison between the posttest score and the cutoff for promotion (i.e., 1=greater than or equal to cutoff, 0=below cutoff).

The program's impact on the likelihood that participants would pass the alternative assessment, and would ultimately be promoted was gauged through a regression analysis that compared the groups' adjusted posttest scores after employing their adjusted pretest scores to account for differences in the adjusted posttest scores. The adjusted posttest scores were converted to pass/fail outcomes based on whether or not the scores met or exceeded the criterion for promotion (i.e., a score of 45th percentile).

Logistic regression, specially designed to analyze problems with pass/fail outcomes was used to conduct the analysis. The predictor weights that result from logistic regression give the influence of the predictors on the outcome expressed as a change in the odds ratio (i.e., the probability of passing divided by the probability of failing). The logistic regression included participation in the program (1=yes, 0=no) and adjusted pretest as predictors in the model. An interaction term, program x adjusted pretest was also included to account for the possibility that the odds of passing were influenced by the students' adjusted pretest scores.

Student Growth

Design

A multiple time series design (Campbell & Stanley, 1963) was used to examine student test administrations (modeled as a within-subjects factor) and to compare the performance of selected subgroups modeled as between-subjects factors. The within-subjects factor represents time (i.e., the spring and summer administrations). The between-subjects factors represent the various levels of the demographic subgroups that were identified as having a significant effect on achievement during the demographic adjustment process described above. This design provided for an examination of the main and interactive effects of the independent variables (i.e., subgroups and time) on the dependent variable (i.e., alternative assessment).

Sample

The sample for this portion of the study was the same as was used in the study of impact except

that a control group was not defined. The analysis included only those students who had valid pre- and posttest scores.

Instrumentation

The posttest for this portion of the study was the SAT-10, reading comprehension subtest. The ITBS-C, reading comprehension subtest served as the pretest. A mathematical procedure was used to apply a difficulty correction to the pretest and transform its scale to make it comparable to the posttest. **This process is an approximation based on limited information. It is not an equating[2] process.** A detailed description of the transformation process may be found in Appendix B.

Data Analysis

A repeated-measures Analysis of Variance (*ANOVA*) was used to apply the multiple time series design. It examined test administrations modeled as a within-subjects factor and selected demographic subgroups modeled as between-subjects factors. This design provided for an examination of the main and interactive effects of the independent variables (i.e., subgroups and time) on the dependent variable (i.e., alternative assessment).

A repeated-measures *ANOVA* creates linear transformations of the within-subjects variables. One variable is created for each degree of freedom of the within-subject factor. An additional variable is created for the average of the repeated-measures. The analysis of variance examines differences in the groups on the transformed variables, rather than on the original within-subject variables (Becker, 1999). This process completely removes within-subjects variability, due to individual differences, from the error term.

As the goal of this analysis was an examination of the equitable distribution of growth, interactions among the various demographic subgroups were not examined. The following dichotomous subgroup classifications were defined: Male, Black, Reduced Price Lunch, English Language Learner, and Special Education.

RESULTS

The evaluation of the Reading Camps was guided by research questions with two foci. The primary focus (questions 1 and 2) was the program's impact on students' alternative assessment scores in terms of improved performance, and in terms of improved likelihood of scoring high enough to be promoted. The second focus (question 3) was the equitable distribution of growth among the various levels of selected demographic subgroups.

Program Impact

The application of the quasi-experimental designs involved a two phase process that used separate data analyses to adjust students' pre- and posttest scores to account for the effects of

---

[2] Equating is a formal statistical process in which a sample of test takers is simultaneously exposed to selected items on two tests, difficulty estimates are generated through Item Response Theory, and the scale of one instrument is mapped onto the scale of the other instrument.

demographic characteristics and then compared the groups' adjusted posttest scores, while controlling for students' initial ability as measured by their adjusted pretest scores.

The first phase used regression analysis to separately estimate the effect of selected demographic characteristics (i.e., gender, ethnicity, reduced price lunch eligibility, ELL status, SPED classification, and age relative to grade) on each student's pretest and posttest scale score in order to account for the influence of those factors when contrasting the groups' performance. Regressions were fitted using a full model and forward selection. Predictors that were not significant ($p < .05$) in either model were removed. The results of this process are displayed in Table 5 lists for each predictor separately for the pre- and posttest, the mean and standard error of its unstandardized regression weight, the standardized weight, and the results of a *t* test that gauges the statistical significance of the weights. The unstandardized weights are expressed in original units (in this case scale scores). Because the pre- and posttest are on different scales the unstandardized weights cannot be compared.

Table 5
Regression Analysis Used for the Demographic Adjustment of the Pre- and Posttest
Predictor Weight Statistics

Predictor	Pretest [a]				Posttest [a]			
	Unstandardized (B)				Unstandardized (B)			
	M	*SE*	Standardized (β)	*t*	*M*	*SE*	Standardized (β)	*t*
Intercept	167.19	0.73		230.56	611.19	1.80		338.65
Special Education	-4.01	0.63	-0.14	-6.36	-17.77	1.44	-0.27	-12.32
Black	-2.47	0.51	-0.11	-4.88	-6.71	1.21	-0.13	-5.55
English Lang. Lear.	--	--	--	--	-5.55	1.44	-0.09	-3.87
Reduced Price Lunch	-3.47	0.77	-0.10	-4.54	-6.38	1.76	-0.08	-3.63
Male	--	--	--	--	-4.12	1.15	-0.08	-3.59

Note. The weights represent the influence on the criterion variable of a unit change in the predictor. Unstandardized weights are in original units. Standardized weights result from rescaling all variables to zero mean and unit variance. The t statistic represents the ratio of the mean weight to its standard error and tests whether the weight is significantly different from zero. All weights are statistically significant ($p < .001$). Statistics for weights that were non-significant ($p \geq .05$) are shown as dashes
[a] Criterion variable

Table 5 shows that on the pretest, students with Special Education (SPED) classifications score (B = 4.01) points lower than those who are not so classified. Black students score 2.47 points lower than non-Black students do, and students who are eligible for the Free/Reduced Price Lunch (FRL) program score 3.47 points lower than those who are not eligible. The results for the posttest are interpreted similarly to those of the pretest. However, the weights are larger due to the larger variance of the scale and additional predictors (i.e., Male and English Language Learner classification) have a significant effect on the outcome.

The standardized weights give the impact of the predictors on the outcome when each is expressed in standard measures (i.e., zero mean and unit variance). This allows the impact of each predictor to be compared and causes the intercept to be fixed at zero. The table shows that SPED classification (β = -0.14) has the greatest influence on the pretest. Both classification as Black (β = -0.11) and FRL eligibility (β = -0.10) have similar but somewhat less influence than SPED classification does. The relative influence of classification as Black and FRL eligibility on

the posttest is similar to that of the pretest. However, the influence of SPED classification is much more pronounced.

An estimate was produced for each student's pre- and posttest score that was subtracted from their actual score and replaced with an average value resulting in adjusted pretest and posttest scores that represented the pretest and posttest scores s/he would have had had s/he had the demographic characteristics of the sample. The results of this process are displayed in Table 6.

Table 6
Results of the Demographic Adjustment Process

| | | Pretest | | | | Posttest | | | |
| | | Unadjusted | | Adjusted | | Unadjusted | | Adjusted | |
Group	*n*	*M*	*SD*	*M*	*SE*	*M*	*SD*	*M*	*SE*
Program	1,825	162.3	11.0	162.4	10.8	596.0	26.0	596.2	24.5
Control	90	162.9	11.3	161.4	10.9	590.0	23.3	585.7	21.4

The table lists for each group, the sample size and descriptive statistics for the unadjusted and adjusted scores, delineated separately for the pre- and posttest. Table 6 shows that the adjustment process tends to inflate the scores of high risk students and shrink the scores of low risk students. Examination of the tabulated data shows that the adjustments served mainly to shrink the posttest scores of the control group. The remaining adjustments were small by comparison.

Status

The question of whether participants scored higher on the alternative assessment than non-participants did was answered by conducting the second phase of the analysis of student outcomes expressed in original scale score units. Regression analysis was used to apply the quasi-experimental non-equivalent control design and to compare the groups' adjusted-posttest scores. It estimated the impact of the program the students' adjusted posttest scale scores while controlling for students' adjusted pretest scores. participation in the program (1=yes, 0=no) and adjusted pretest as predictors in the model. An interaction term program x adjusted pretest score was also included to account for the possibility that the impact of the program was influenced by the students' adjusted pretest scores.

Table 7 displays the results of this process. It lists for each predictor; descriptive statistics (i.e., mean and standard error) for its unstandardized regression weight, the standardized weight, and the results of a *t* test that gauges the statistical significance of the weights.

Table 7 lists in its second and third columns the unstandardized weights, which appear in original scale score units. The table shows that a one unit change in the (adjusted) pretest above the sample mean produced a 0.98 unit change in the (adjusted) posttest. Students who participated in the program scored 9.45 points higher than those who did not. The non-significant program x pretest interaction indicates that the effect of the program was not affected by the students' pretest scores. Although, comparison of the standardized weights shows that the pretest had a much greater influence on the posttest than did the program (i.e., where students started was the

largest predictor of where they finished), students tended to gain 9.5 point regardless of their pretest score.

Table 7
Regression Analysis of the Adjusted Scores

| | Unstandardized | | | |
Effects	*M*	*SE*	Standardized	*t*
Intercept	595.70	0.50		1185.29
Pretest	0.98	0.05	0.43	20.95
Program	9.45	2.38	0.08	3.98
Program x Pretest	--	--	--	--

Note. The weights represent the influence on the criterion variable of a unit change in the predictor. Unstandardized weights are in original units. Standardized weights result from rescaling all variables to zero mean and unit variance. The t statistic represents the ratio of the mean weight to its standard error and tests whether the weight is significantly different from zero. All weights are statistically significant ($p < .001$). Statistics for weights that were non-significant ($p \geq .05$) are shown as dashes. The pre- and posttest scores are adjusted for demographics.

The application of a correction to the adjusted posttest using the same process that was used in the demographic adjustment was used to generate a second adjustment to account for the effects of the adjusted pretest. This permitted the groups' mean scores to be compared directly.

The results of this process are displayed in Table 8. It lists for each group, the sample size and descriptive statistics for the unadjusted and adjusted scores, delineated separately for the pre- and posttest.

Table 8
Results of the Pretest Adjustment

| | | Unadjusted | | Adjusted | |
Group	*N*	*M*	*SD*	*M*	*SD*
Program	1,825	596.0	26.0	596.1	22.6
Control	90	590.0	23.3	586.7	20.5

A comparison of the adjusted means in Table 8 using a *t* test for independent samples produce results that are identical to those shown in the regression analysis. The effect size ($d = .43$) of the program's impact derived from those *t* tests, has a weak-moderate level of practical significance based on Cohen's classification.[3]

Promotion

The program's impact on the likelihood that participants would pass the alternative assessment, and would ultimately be promoted was gauged through a regression analysis that compared the groups' adjusted posttest scores after their adjusted pretest scores were taken into account. The adjusted posttest scores were converted to pass/fail outcomes based on whether or not the scores met or exceeded the criterion for promotion (i.e., a score of 45th percentile).

Logistic regression, specially designed to analyze problems with pass/fail outcomes was used to conduct the analysis. The predictor weights that result from logistic regression give the influence

---

[3] Cohen (1988) has classified the difference between two means as .20 (weak), .50 (moderate), and .80 (strong).

of the predictors on the outcome expressed as a change in the odds ratio (i.e., the proportion that pass divided by the proportion that fail). The analysis included the program (1=yes, 0=no) and adjusted pretest score as predictors in the model. An interaction term, program x adjusted pretest score was also included to account for the possibility that the odds of passing were influenced by the students' adjusted pretest scores. The conditional probability of attaining the passing standard given their adjusted pretest score was compared for the students in the Reading Camps and the students in the control group.

Table 9 lists for each effect, the predictor weight (*B*) and its standard error (*SE*) followed by the change in the odds ratio due to the predictor, and the Wald statistic (a measure of the statistical significance of the effect given by the square of the ratio of *B* to its standard error).

Table 9
Logistic Regression Analysis Adjusted Posttest Scores

Effect	*B*	S.E.	Odds Ratio	Wald
Intercept	-1.76	0.07	0.17	584.25***
Pretest	0.08	0.01	1.09	145.59***
Program	1.37	0.47	3.92	8.35**
Program x Pretest	--	--	--	--

Note. All predictors are grand-mean centered. *B* gives the predictor weight, which is the influence of the indicated effect on the outcome variable expressed on a log-odds scale. Wald is a measure of statistical significance given by the square of the ratio of *B* to its standard error. The odds ratio is found by raising the base of the natural logarithm to the power given by the predictor. This gives the probability of meeting the criterion divided by the probability of not meeting the criterion. Variables are adjusted scores.
** $p < .01$. *** $p < .001$.

The intercept term gives the odds of passing the alternative assessment for the typical student, which were .17 to 1. This equates to an adjusted pass rate (odds/1+odds) of 14.5%. Each one point increase in the pretest score above the sample mean improves those odds by a factor of 1.09 to 1. The odds of passing for students who completed the Reading Camps relative to those who did not were 3.92 to 1. The non-significant interaction indicates that those odds were significantly affected by the pretest. The practical significance of the program's impact as determined through direct examination of the odds ratio for program participation (OR = 3.92) was moderate to strong, based on Chinn's formula and Cohen's classification.[4]

To summarize, analysis of the Reading Camps' impact on students' alternative assessment scores showed that students who completed the program scored 9.5 scale score points higher than students who did not participate in the program, when students' pretest scores and demographic characteristics were taken into account. This difference in the groups' means had weak practical significance. However, when the data was reanalyzed to determine the program's impact on the attainment of the alternative assessment score required for promotion, it was shown that students who completed the program had 3.9 greater odds of meeting the criterion than students who did not participate in the program, when students' pretest scores and demographic characteristics were taken into account. That difference in the groups' odds had moderate-strong practical significance.

---

[4] Chinn (2000) developed a formula to convert d to an odds ratio for use in meta-analysis. The breakpoints obtained by applying this formula to Cohen's (1988) classification are 1.5 (weak), 2.5 (moderate), and 4.3 (strong).

Student Growth

This section of the evaluation addressed the growth that participating students achieved as a result of attending summer school. It examined performance improvement from the spring to the summer administration of the alternative assessment. The equitable distribution of growth among the levels of selected demographic subgroups was also explored.

A multiple time series design was used to examine student test administrations modeled as a within-subjects factor – Time - (i.e., the spring and summer administrations), and to compare the performance of selected subgroups modeled as between-subjects factors. The sample and instrumentation for this portion of the study was the same as was used in the study of program impact, but a control group was not included. A mathematical procedure was used to apply a difficulty correction to the pretest and transform its scale to make it comparable to the posttest.

Descriptive statistics for each subgroup of the pre- and posttest are presented in Table 10. The table lists for each subgroup, the sample size, the mean and standard deviation of the pretest, and the mean and the standard deviation of the posttest.

Table 10
Alternative Assessments Scores of Students who Completed the Reading Camps
By Demographic Subgroup

Group	n	Pretest M	Pretest SD	Pretest M	Pretest SD	Gain M	Gain SD
Sex							
Male	1,092	555.3	28.2	593.8	26.9	38.4	28.0
Female	733	557.2	27.6	599.2	24.3	42.0	27.7
Difference[a]	1,825	-1.9	(1.4)	-5.4	(1.2)	-3.6	(1.4)
Race							
Black	789	552.4	27.8	593.0	25.0	40.6	28.9
non-Black	1,036	558.9	27.8	598.2	26.5	39.3	27.2
Difference[a]	1,825	-6.5	(1.3)	-5.2	(1.2)	1.3	(1.3)
Free/Reduced Lunch							
FRL	1,628	555.0	27.7	594.9	25.7	40.0	28.0
non-FRL	197	565.2	28.9	604.5	27.0	39.3	27.0
Difference[a]	1,825	-10.2	(2.1)	-9.6	(1.9)	0.7	(2.1)
English Lang. Lear.							
ELL	402	556.6	27.5	592.6	25.5	36.1	26.0
non-ELL	1,423	556.0	28.1	596.9	26.1	41.0	28.4
Difference[a]	1,825	0.6	(1.6)	-4.3	(1.5)	-4.9	(1.6)
Special Education							
SPED	351	548.3	29.2	581.4	27.8	33.1	30.9
non-SPED	1,474	557.9	27.4	599.4	24.3	41.5	26.9
Difference[a]	1,825	-9.6	(1.6)	-18.0	(1.5)	-8.4	(1.6)
Overall	1,825	556.1	28.0	596.0	26.0	39.9	27.9

Note. Pretest scores are corrected for difficulty and transformed to the scale of the posttest
[a] Differences between the subgroup levels represent the top minus the bottom levels. Values enclosed in parenthesis represent standard errors.

Table 10 shows that there were achievement gaps within the levels of the various subgroups. The pretest mean ranged from a low of 548.3 (SPED) to a high of 565.2 (non-FRL) and averaged 556.09. The posttest mean was much higher. It ranged from a low of 581.4 (SPED) to a high of

604.5 (non-FRL) and averaged 596.0. Gains from pre- to posttest were quite large, though some variation was seen. The gains ranged from a low of 33.1 (SPED) to a high of 42.0 (Female) and averaged 39.9. Differences in the gains were also seen between the levels of the subgroups. These were -8.4 (SPED vs. non-SPED) followed by -4.9 (ELL vs. non-ELL) followed by -3.6 (Female vs. Male) with advantages seen for (Black vs. non-Black) and (FRL vs. non-FRL).

A repeated-measures Analysis of Variance (*ANOVA*) was used to apply the multiple time series design and test the statistical significance of the differences seen. It examined the main effect of the within subjects factor and of the between subjects factors (Sex, Race, Reduced Price Lunch, English Language Learner, and Special Education) and the interaction between the within subjects main effects and the between subjects main effects.

Table 11 summarizes the results of the repeated measures *ANOVA* of the groups' alternative assessment scores. The table lists for each source of variation, the F statistic used to gauge its statistical significance, its effect size η,[5] and its significance level *p*)

Table 11
Repeated Measures Analysis of Variance:
Effect of Selected Demographic Factors on Students' Alternative Assessment Scores

Source	df	F	η	p.
Between subjects				
Sex	1	5.26	.05	.02
Race	1	37.65	.14	.00
Reduced Price Lunch (FRL)	1	22.43	.11	.00
English Language Learner (ELL)	1	6.79	.06	.01
Special Education (SPED)	1	114.06	.25	.00
S within-group error	1,819	(967.20)		
Within subjects				
Time	1	725.96	.53	.00
Time x Sex	1	4.49	.04	.03
Time x Race	1	0.09	.01	.76
Time x FRL	1	0.30	.01	.59
Time x ELL	1	8.73	.06	.00
Time x (SPED)	1	23.21	.09	.00
Time x S within-group error	1,819	(381.77)		

Note. Values enclosed in parenthesis represent mean squared errors. *S* = subjects. Between-subjects effects are based on the sum of the across time. Within-subjects effects are based on the difference in the scores over time. F is a statistic used to gauge the ratio between the error due to an effect and within group error. η is a measure of effect size.

The between-subjects main effects describe differences in the between-subjects factor (i.e., the average of the pre- and posttest score) for the subjects at the different levels (i.e., ELL vs. non-ELL) of the subgroup, while the within-subjects main effect of time describes the average change in the scores from pre- to posttest. The within-subjects interaction terms (e.g., Time x Sex) represent systematic differences in the growth (i.e., pre- to posttest gain) experienced by the

---

[5] Cohen (1988) has classified the practical significance of the effect size η as .10 (weak), .24 (moderate), and .62 (strong).

subjects at the different levels (i.e., Male vs. Female) of the subgroup.

Table 11 shows that significant between-subjects main effects were identified for each of the subgroups. This indicates that there were differences in the between-subjects factor within each of the identified subgroups. Examination of the effect size shows that the differences were largest for SPED (moderate), followed by Race and Reduced Price Lunch (weak). These differences confirm what was found through examination of the means.

Table 11 also shows a significant and strong effect for time. This indicates that substantial overall student growth was experienced from pre- to posttest. These results are also confirmed by examination of the means. Though caution must be exercised in quantifying this growth in light of the fact that the scores were not formally equated, the effect sizes seen are similar to those found by comparing the scores of students who were retained in third grade on the two most recent consecutive administrations of the FCAT-NRT (Urdegar, 2009). The table also shows significant within-subject interactions for Race, ELL, and SPED. Those differences had effect sizes with weak levels of practical significance. However, the difference for SPED vs. non-SPED students had an effect size with nearly a moderate level of practical significance.

In sum, analyses examining the distribution of the status and growth in students' achievement scores across selected demographic subgroups revealed significant achievement gaps for students who were male, black, FRL eligible, ELL, and placed in SPED that mirrored the results of other studies (e.g., Eamon, 2005; Levitt et al. 2005; McLoyd, 1990). Though strong growth was seen for participating students as a whole, the gains of students who were designated as ELL and SPED was significantly lower than students who were not so designated.

DISCUSSION

The fundamental goal of the Reading Camps program is to bolster the students' reading skills and thus prepare them for the next opportunity to pass an alternative assessment that would enable third grade students who were scheduled for retention to be promoted to the fourth grade. Two alternative assessments were offered to students. The Iowa Test of Basic Skills, Edition C (ITBS-C) was administered at the end of the regular school year and the Stanford Achievement Test, Tenth Edition (SAT-10) was administered at the end of summer school.

The district undertook an evaluation to gauge whether students who completed the program exhibited improved performance on the alternative assessment when compared to a comparison group of students who did not participate. Also examined was students' spring to summer growth on the alternative assessments and the extent to which that growth was influenced by participants' demographic characteristics.

Summary

The evaluation of the Reading Camps described the operation and assessed multiple measures of outcome for the program that operated during its seventh year of implementation. Conducted by the district's Office of Program Evaluation, the study was guided by a series of questions that can now be answered.

1. **Did students who participated in the Reading Camps score higher on the summer alternative assessment than students who did not participate when their spring alternative assessment scores and demographic factors were taken into account?**

   The findings demonstrate that students who completed the program scored 9.5 points higher than students who did not participate in the program, when students' pretest scores and demographic characteristics were taken into account. This gauge of the program's impact had weak-moderate practical significance. Therefore, the first question was answered affirmatively.

2. **Were students who participated in the Reading Camps more likely to score high enough on the summer alternative assessment to be promoted than students who did not participate, when their spring alternative assessment scores and demographic factors were taken into account?**

   The results show that students who completed the program had 3.9 greater odds of meeting the cutoff for promotion than students who did not participate in the program, when students' pretest scores and demographic characteristics were taken into account. This gauge of the program's impact had moderate-strong practical significance. Therefore, the second question was answered affirmatively.

3. **Did students who attended the Reading Camps show growth on the alternative assessment from spring to summer, and how was that growth influenced by the students' demographic characteristics?**

   The results show significant achievement gaps for students who were male, black, FRL eligible, ELL, and placed in SPED that mirrored the results of other studies (e.g., Eamon, 2005; Levitt et al. 2005; McLoyd, 1990). Though strong growth was seen for participating students as a whole, the gains of students who were designated as ELL and SPED was significantly lower than students who were not so designated. Therefore, the third question was answered affirmatively for all student subgroups except ELL and SPED.

Conclusions

The program has consistently yielded favorable results since its inception. The Reading Camp program has for the seventh time proven to be an effective means of bolstering both the students' reading skills and their chances of passing the alternative assessment test. Though strong growth was seen for participating students as a whole, the achievement gaps and attenuated growth of students who were designated as ELL and SPED, relative to students who were not so designated, suggests that additional strategies be developed to benefit those students. The small percentage of non-participants with posttest data limits the generalizability of this study, as the performance of the few students who were included may not be representative.

REFERENCES

Becker, L.A. (1999). GLM repeated measures: One within, one between. *SPSS Course Content – Fall 2000*. (Web based syllabus). University of Colorado at Colorado Springs. Retrieved August 26, 2009, from http://web.uccs.edu/lbecker/SPSS/glm_1w1b.htm

Campbell, D.F., &, Stanley, J.C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine, 19* (22). 3127-3131. Retrieved Aug 20, 2009, from http://www3.interscience.wiley.com/cgi-bin/fulltext/75500445/PDFSTART

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

Department of Summer Services (2009). Guidelines, procedures, and program information [Electronic version]. In 20*09 educational summer services, 2009 summer implementation document* (Ch. 3, pp. 9 – 18). Miami, FL: Miami-Dade County Public Schools. Retrieved July 22, 2009, from http://summerschool.dadeschools.net/pdfs09/ch_3.pdf

Eamon, M. K. (2005). Social-demographic, school, neighborhood, and parenting influences on the academic achievement of Latino young adolescents [Electronic version]. *Journal of Youth and Adolescence, 34*(2), 163-174.

Florida Department of Education. (2005). *Critical information for parents of third graders* [Electronic version] Retrieved July 23, 2009 from http://fcat.fldoe.org/mediapacket/2005/pdf/05gr3ReadtoLearn.pdf

Gomez, J.J. (2008). *Evaluation of the third grade summer reading camps in the summer of 2008*. Miami, FL: Miami-Dade County Public Schools.

Levitt, J. L., Shay, S. A., Urdegar, S. M., Shpurik, M., & Sorhaindo, L. R. (2005). *Title I evaluation summary report, 2003-04*. Miami, FL: Miami-Dade County Public Schools.

McLoyd, V. C. (1990). The impact of economic hardship on black families and children: Psychological stress, parenting, and socioemotional development. *Child Development, 61*(1).

Public School Student Progression; Remedial Instruction; Reporting Requirements [Electronic version], 48 Fla. Stat. §1008.25 (2008). Retrieved July 22, 2009, from http://www.leg.state.fl.us/STATUTES/index.cfm?App_mode=Display_Statute&URL=Ch1008/titl1008.htm&StatuteYear=2008&Title=%2D%3E2008%2D%3EChapter%201008

Urdegar, S.M. (2009). *A within group analysis of retained third graders' performance on the FCAT-NRT*. Unpublished data analysis. Miami, FL: Miami-Dade County Public Schools.

Appendix A

List of Summer Reading Camps

Table A
List of Summer Reading Camps

Area	Work Location	School Name
N	0641	Bunche Park Elementary
	0761	Fienberg/Fisher K-8 Center
	3421	Marcus A. Milam K-8 Center
	3861	North Glade Elementary
	4001	Norwood Elementary
NC	1521	Amelia Earhart Elementary
	0101	Arcola Lake Elementary
	5021	Ben Sheppard Elementary
	2361	Hialeah Elementary
	0081	Lenora Braynon Smith
	2981	Liberty City Elementary
	2911	Linda Lentin K-8 Center
	3431	Phyllis R. Miller Elementary
SC	1081	Coral Terrace Elementary
	1811	Dante B. Fascell Elementary
	1841	Flagami Elementary
	5561	Frances S. Tucker Elementary
	1361	Frederick Douglass Elementary
	1881	H. M. Flagler Elementary
	4741	Royal Green Elementary
	5001	Shenandoah Elementary
	5521	Tropical Elementary
S	2321	Gulfstream Elementary
	2901	Leisure City K-8 Cntr
	3261	Miami Heights Elementary
	4441	Pine Lake Elementary
	3541	Robert Russa Moton Elementary
	5791	West Homestead Elementary

Appendix B

Scale Transformation

The correction and rescaling of the pretest (i.e., ITBS-C) to be comparable to the posttest (i.e., SAT-10) consisted of a three step process that involved (a) estimation of the scale distribution for the pre- and posttest, (b) computation and application of a difficulty correction, and (c) rescaling. First, the standard deviation of the scale scores of the pretest was estimated by subtracting the scale score corresponding to the top sixth of the percentile scale from the scale score corresponding to the bottom sixth of the percentile scale as interpolated from the published norms and dividing by the appropriate normal deviate (i.e., 1.935 standard deviations).

Then, the probability of scoring at or less than each scale score of the pretest was computed. These probabilities were then converted to odds ratios (i.e., probability/ 1- probability). The log of each odds ratio was then taken. Log odds ratios represent equal units of difficulty that are used in Item Response Theory to develop equally spaced scales of measurement.

The different pass criteria for the pretest (50th percentile) and the posttest (45th percentile) suggest that the posttest is a more difficult test. Therefore, a difficulty correction equal to the log odds (45th percentile) was applied to the log odds ratios of each scale score of the pretest. The corrected log odds ratios were then converted back to odds ratios, which were converted to probabilities (i.e., odds/1+odds). The probabilities were then converted into standard normal deviates and transformed to the scale of the posttest.