Miami-Dade County Public Schools

Third Grade Summer Reading Camps, 2012 Evaluation

January 2013

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

ASSESSMENT, RESEARCH AND DATA ANALYSIS
1450 Northeast Second Avenue
Miami, Florida 33132

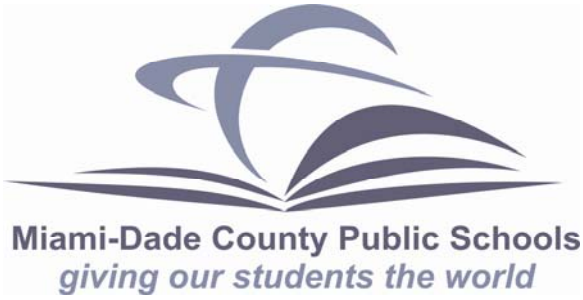
Sally A. Shay,
District Director

THE SCHOOL BOARD OF MIAMI-DADE COUNTY, FLORIDA

Ms. Perla Tabares Hantman, Chair
Dr. Lawrence S. Feldman, Vice Chair
Ms. Susie V. Castillo
Mr. Carlos L. Curbelo
Dr. Wilbert "Tee" Holloway
Dr. Martin Karp
Dr. Dorothy Bendross-Mindingall
Dr. Marta Pérez
Ms. Raquel A. Regalado

Mr. Jude Bruno, Student Advisor

Mr. Alberto M. Carvalho
Superintendent of Schools



Gisela F. Feild
Administrative Director
Assessment, Research and Data Analysis

TABLE OF CONTENTS

LIST OF TABLES

EXECUTIVE SUMMARY

The fundamental goal of the Third Grade Summer Reading Camp program is to bolster the reading skills of third grade students scheduled for retention and to prepare them to demonstrate mastery of state standards in order to be promoted to the fourth grade. The Office of Program Evaluation undertook an evaluation to gauge whether students who completed the program in Summer 2012 exhibited improved performance on the summer alternative assessment when compared to a comparison group of students non-participants. The summer alternative assessment, the Stanford Achievement Test, Tenth Edition (SAT-10), was offered to students after the reading camp and served as the posttest. The spring alternative assessment, the Iowa Test of Basic Skills, Edition C (ITBS-C), served as the pretest.

Principals were surveyed regarding their opinions of the reading camps and generally gave positive accounts of planning, transportation, and curriculum/instruction. However, considerable variation between the schools in the duration of the instruction provided to the students using *After the Bell* was found. Therefore, the program could not be judged to be consistently implemented.

Although, a sizable majority of principals endorsed the level of technical support provided by District reading specialists, all felt that the program was not effective at improving the students' reading skills. Nonetheless, more than half agreed that the program should continue in its current form.

Students who completed the program did not score significantly higher than those who did not participate when their pretest scores and demographic characteristics were taken into account. Moreover, students who completed the program did not show a greater chance of meeting the cutoff for promotion than those who did not participate. Strong growth from pre- to post- test was seen for participating students as a whole, but differences in the levels of the Special Education (SPED) demographic subgroup were less pronounced.

INTRODUCTION

The Third Grade Summer Reading Camps is an intensive reading remediation program designed to prepare the students to pass the Florida Comprehensive Assessment Test (FCAT 2.0). Students targeted for enrollment are third graders who failed to score above Level 1 on the reading subtest of the FCAT 2.0 and did not qualify for one or more of the statutory "good cause exemptions," and as such are typically prevented from advancing to grade 4. The following report details the evaluation of the tenth implementation of the camps (summer 2012) and focuses on their impact on the students' subsequent performance on the alternative assessment test.

Background

Florida Statutes currently prohibit social promotion and mandate that students with academic deficiencies be provided with intensive remediation with particular emphasis on the reading proficiency of students in the grades K through 3. Districts are 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, . . . , 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*, 2012).

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 or above on the reading subtest of the FCAT 2.0, administered in grade 3. Scoring at Level 1 on the reading subtest of the FCAT may lead to retention in grade 3, unless the student meets one of the following six types of "good cause" exemptions:

- an acceptable level of performance on an alternative assessment;
- demonstration, through a 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; or,
- receipt of two years of intensive remediation in reading, a continuing deficiency in reading, and two previous retentions in grades K through 3.

Because large numbers of the state's third graders have historically scored at level 1, the State called for districts to provide "intensive reading camps" during the summer, for such students designed to prepare the students for the next opportunity to pass the test (Florida Department of Education, 2008).

The program has undergone numerous changes in format and length of delivery since its inception and has produced mixed results. While the results of the most recent evaluation did

show improvement in participating students' reading skills, their chances of passing the alternative assessment were not significantly different from that of non-participants.

Description of the Program

The Reading Camp curriculum that operated this past summer represented a change from the curriculum that had operated for the previous four years. It focused entirely on bolstering students' reading skills and did not contain dedicated test preparation activities. The program in place provided students with three hours of reading-related activities per day, five days per week, for four weeks. It consisted of a single component *After the Bell* that operated within the school day.

A half-day training was provided to teachers by the vendor. Teachers were given the scope and sequence of instruction, previewed materials, and given a suggested schedule for the 3 hour block of instruction. Teachers were given the opportunity to walk through the components of the program and ask questions regarding implementation (Department of Language Arts/Reading, 2012).

The research-based intervention program utilized both whole group and small group instruction. Targeted skills were taught using a program that provided fiction and non-fiction libraries. Each unit targeted reading comprehension skills, phonics instruction, and fluency practice. The following materials were included in the program:

- Student workbooks
- Reading Skills Cards
- Fluency cards
- Phonics reproducibles
- Teacher guides
- Assessment book
- Read-Aloud books
- Independent Reading Library
- Instructional Reading Library

Students were also directed to independent reading activities during small group instruction. Successmaker may also have been used as a supplemental technology component during small group instruction (Department of Language Arts/Reading).

The Department of Language Arts/Reading periodically monitored the delivery of the curriculum during single on-site visits to selected Reading Camps. The program vendor also visited every summer service site at least once. Reports of the findings from the vendor's visits were provided to district administrators (Department of Language Arts/Reading).

The program that was implemented in 2012 provided students with 20 half-days of instruction prior to the alternative assessment. The District's, *Implementation Documentation: 2012 Summer Learning Waves of Learning* (Department of Summer Services, 2012) limited participation in the Reading Camps to "students who were retained in grade 3" (p. 10). Consequently, the students targeted by the program remained primarily the third graders who scored at Level 1 on the reading portion of the FCAT 2.0 administered in April of 2012 and did not meet one of the "good cause" exemptions outlined earlier in this report.

Table 1
Features of the Third Grade Summer Reading Camps

Program Features

	2008	2009	2010	2011	2012
Hours per day [a]	6	3	3	3	3
Days per week	5	5	5	5	5
Daily schedule	8:30-2:30	8:30-11:30 [b]	8:30-11:30 [b]	8:30-11:30 [b]	8:30-11:30 [b]
Weeks per term	4	5	5	5	5
Hours of instruction per term	110	60	60 [c]	60	60
Teacher-student ratio	1:18	1:18	1:18	1:18	1:18
Number of camps	42	28	26	33	34
Student enrollment	2,466	2,116	2,375	2,273	1,833

Note. Each column represents the schedule in effect for that year.

[a] The hours per day includes ½ hour for lunch. [b] Wednesday schedule was 8:30-2:00. [c] Post testing was conducted after 17 days of instruction

Table 1 provides information on various operational features of the Reading Camp program in each summer that it was implemented during the last five years. One sees that the greatest number of changes, which occurred from 2008 to 2009, decreased the length of the instructional day by nearly half, decreased enrollment by one-fourth, and decreased the number of sites by one-third.

In 2010 students received four fewer days of instruction prior to receiving the summer alternative assessment. The present implementation provides students with the full program prior to the administration of the assessment. The schools that implemented Reading Camps in 2012 are listed in Table 2.

Table 2
List of the Reading Camps, 2012

Region I		Region II	
Location	School Name	Location	School Name
0101	Arcola Lake Elementary	0361	Biscayne Gardens Elementary
3781	Barbara Hawkins Elementary	5005	David Lawrence K-8 Center
5991	Charles D. Wyche Elementary	0761	Fienberg/Fisher K-8 Center
1161	Crestview Elementary	1481	John G. Dupuis Elementary
2281	Greynolds Park Elementary	2911	Linda Lentin K-8 Center
2361	Hialeah Elementary	3421	Marcus A. Milam K-8 Center
1681	Lillie C. Evans K-8 Center	5971	Nathan B. Young Elementary
0111	Maya Angelou Elementary	4261	Palm Springs Elementary School
		3431	Phyllis R. Miller Elementary
		2371	West Hialeah Gardens Elementary

Region IV		Region V	
Location	School Name	Location	School Name
0451	Ashe/Doolin K-8 Academy	0261	Bel-Aire Elementary
0121	Auburndale Elementary	0651	Campbell Drive K-8 Center
1001	Coral Park Elementary	0861	Colonial Drive Elementary [a]
1121	Coral Way K-8 Center	4651	Ethel F. Beckford/Richmond Elementary [b]
1841	Flagami Elementary	2321	Gulfstream Elementary
5561	Frances S. Tucker Elementary	4391	Irving & Beatrice Peskoe K-8 Center
4761	Royal Palm Elementary	2651	Kendale Lakes Elementary [c]
		2941	Laura C. Saunders Elementary
		3261	Miami Heights Elementary
		5421	Sunset Park Elementary

Note. Camps with superscripts were physically housed at the elementary schools indicated.
[a] Gloria Floyd. [b] Coral Reef. [c] Oliver Hoover.

METHODOLOGY

The Reading Camps are designed to provide intensive reading instruction to third graders, slated for retention due to insufficient reading skills. The program aims to prepare them to score high enough on an alternative standardized reading assessment to be promoted to fourth grade.

Research Questions

An evaluation was undertaken by the district's Office of Program Evaluation to assess the implementation of the program, to explore the attitudes of responding principals toward the program, and to gauge its impact. The evaluation was guided by a series of questions:

1. **Were the Reading Camps implemented as planned?**
2. **Did students who participated in the Reading Camps score higher on the summer alternative assessment than students who did not participate?**
3. **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?**
4. **Did students who attended the Reading Camps show growth in their reading skills from spring to summer?**

Data Sources

Data were gathered from four sources to address the evaluation questions. The first source of data 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 second source of data was an online instrument designed to measure various aspects of the summer program including enrollment, planning, staffing, transportation, resources, materials, supplies, and specific curricular-factors. The third source and fourth source of data were students' demographic and assessment records maintained on the District's mainframe computer system. Each of the data sources will be discussed in detail in ensuing sections.

Implementation

Implementation was examined so that any observed effects on participants' achievement could be properly attributed to the program. This portion of the evaluation was concerned with determining whether the Reading Camps were being operated as specified; and, whether or not sufficient planning, resources, and support were evident. Data were gathered from an online survey entitled the Summer School Implementation Survey to gauge implementation. (A copy of the survey may be found in Appendix A). The sample of schools for this portion of the evaluation included all participating schools.

The Summer School Implementation Survey measures various aspects of the summer program including enrollment, planning, staffing, transportation, resources, materials, supplies, and specific curricular-factors (i.e.; dosage, frequency, and duration of treatment; engagement; and, perceived effectiveness). The survey comprised 33 items of which 24 adhered to a Likert-type

format with response options that varied. Two filter questions (16 and 22) were used to enable a branching process by which only an appropriate subset of items were presented to respondents. Respondents were typically exposed to 24 items and two filter questions. Five items (1, 4, 5, 23 and 24) were used for identification, programmatic information, and to provide space for schools that did not offer the district-defined curriculum to describe their program. Two items (2 and 3) gauged enrollment, six items (5-10) assessed the adequacy of planning, resources, and staffing levels, and three items (11 – 13) pertained to transportation.

Seventeen program-specific curricular-questions were also defined: Four items (17, 18, 25, and 26) measured dosage, two items (21 and 29) pertained to program effectiveness, and two items (19 and 27) gauged the program's ability to provide differentiated instruction. Two additional items (20 and 28) measured students' engagement. Three summary items (30-32) solicited respondents' overall impressions of the program and the technical support that was available. Finally, space was provided to give respondents the option of suggesting areas for improvement. In all, 17 items measured implementation. The survey was administered online to principals of participating sites during July and August of the 2012-13 school year. Key items were organized into one of four implementation categories: Planning, Resources, Transportation, and Curriculum/Instruction. The percent of positive responses within a category are classified as 0-50 (not implemented), 51-69 (partially implemented), and 70-100 (fully implemented. Otherwise, the analysis of the results of the Summer Implementation Survey, was limited to descriptive statistics.

An attitudinal component measured principals' summative perceptions of the program. The issues addressed included technical support, effectiveness, satisfaction with the status quo, and suggestions for improvement. Data for this component were drawn from the Summer Implementation Survey. Three of those items (30-32) measured attitude. The analysis of the results for the attitudinal component was limited to descriptive statistics.

Program Impact

Design and Samples

A non-equivalent control group quasi-experimental design (Campbell & Stanley, 1963) was used to compare the performance of a group of students' who participated in the Reading Camps with a group of students who did not, using pretest scores to equate the groups' performance prior to exposure and posttest scores to measure their performance afterwards. The groups were considered nonequivalent, because group membership was assigned randomly.

The population for this evaluation consisted of third grade students who were scheduled for retention after having exhausted all options for promotion available to them prior to the end of the regular school year. Of the 27,204 third grade students who took the reading subtest of the FCAT 2.0, and were active in the district at the end of the 2011-12 school year, 21.2% (n = 5,776) scored within achievement Level 1. Of those students, 90.2% (n=5,211) faced mandatory retention under state statute. A subset qualified for one or more good cause exemptions outlined by the state. Table 2 lists the number and percent of students who utilized those exemptions.

Students listed as receiving exemptions based on the Portfolio assessments earned a passing score on that test. The remaining 2,304 students were eligible to attend the Reading Camps. The students who elected to attend and went on to complete the Reading Camps were included in the treatment group.

Table 3
Good Cause Exemptions to Mandatory Retention Granted to M-DCPS Third Graders Who Scored at Level 1 on the FCAT Reading Subtest, 2011-12

	n	%
English Language Learner [a]	730	25.1
Portfolio Assessment	1,297	44.6
Alternative Assessment [b]	212	7.3
Students with Disabilities [c]	556	19.1
Other [d]	112	3.9
Total	2,907	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] Includes ITBS-C given in spring. [c] Students with Disabilities who were retained once and had two years of intensive remediation. [d] Students with Disabilities for whom statewide assessment is not appropriate; and, students who received intensive remediation for two or more years and were previously retained for two years.

A comparison group was also defined which consisted of the eligible students 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 4 partitions the students slated for retention according to their current and prior participation in the Reading Camps.

Table 4
Crosstabulation of Previous vs. Current Program Exposure Status: Students Slated for Retention

	Current Exposure			
Previous Exposure	Complete	Partial	None	Total
No	1,577	22	555	2,154
Yes	100	1	49	150
Total	1,677	23	604	2,304

Table 4 shows that of the 2,174 students slated for retention, a total of 2,154 had no previous exposure to the Reading Camps. The treatment pool was comprised of 1,529 of the 1,577 students who completed the program and did not attend a charter school during the summer. The control pool was comprised of the 555 students who had no current or previous exposure to the program. The other groups were excluded. Of the 1,529 students in the treatment pool, 93.1% (n = 1,424) had valid pre- and post- test (i.e., spring and summer alternative assessment) scores. These students constituted the treatment group. Of the 555 students in the control pool, only 13.5% (n = 75) had valid pre- and post- test scores. These students constituted the control group. The characteristics of the final sample are described in Table 5. It lists for each subgroup, the percentage of students in the treatment and control group.

Table 5

The Groups' Demographic Characteristics as a Percentage of the Sample

Subgroup	Treatment (n = 1,424)	Control (n = 75)
Gender		
Male	61.4	61.3
Female	38.6	38.7
Ethnicity		
Black	41.8	30.7
Non-Black	58.2	69.3
Free/Reduced Price Lunch		
Eligible	94.6	90.7
non-Eligible	5.4	9.3
English Language Learner		
Current	50.8	50.7
Former/Never	49.2	49.3
Special Education		
Disabled	21.6	29.3
non-Disabled	78.4	70.7
Age	9.5	9.4

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

Table 5 shows that the treatment and control groups appear to be similar when compared within each of the demographic subgroups, other than ethnicity. However, none of those differences were found to be statistically significant.

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 for Third Grade Promotion (AATGP) 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 also a standardized norm-referenced test. It is currently administered by the District to third grade students as an alternative assessment at the end of the school year.

Data Analyses

Demographic and pretest differences are known to influence achievement so that any between group variation on such variables can mask the program's impact. Therefore, the application of the quasi-experimental design used separate data analyses to adjust students' pre- and posttest scores to account for the effects of demographic characteristics and then compared the groups' posttest scores, while controlling for students' initial ability. A detailed description of the adjustment process may be found in Appendix B.

Status

Regression analysis was used to apply the non-equivalent control group design and to compare the groups' posttest scores. It estimated the impact of the program on the students' posttest scale scores, while controlling for students' pretest scores. Pretest and Program, 1 (participant) and 0 (non-participant) were the main predictors in the model. An interaction term (Program x Pretest) was also included to determine if the impact of the program differed with students' pretest scores.

Promotion

The question of whether participants were more likely than non-participants to have scored high enough on the alternative assessment to be promoted was also gauged through a regression analysis that estimated the impact of the program on the students' posttest scores, while controlling for students' pretest scores. However, the outcome of interest was whether or not the students posttest exceeded the 45th percentile, which is the cutoff for promotion: 1 (45th - 99th percentile) and 0 (1st - 44th percentile).

Logistic regression, specially designed to analyze problems with pass/fail outcomes was used to conduct the analysis. Pretest and Program, 1 (participant) and 0 (non-participant) were the main predictors in the model, and an interaction term (program x pretest) was included to determine if the impact of the program differed with students' pretests. Demographic characteristics were also included because cutoff scores cannot be adjusted to account for them.

Student Growth

Design and Samples

A multiple time series design (Campbell & Stanley, 1963) was used to compare the performance of the levels of each of several demographic subgroups across administrations of the pre- and post- tests. This design provided for an examination of the main and interactive effects of subgroup and time on the posttest score. The sample for this portion of the study was the same as was used in the study of impact except that no comparison group was 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[1] process.** A detailed description of the transformation process may be found in Appendix C.

[1] 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.

Data Analysis

A repeated-measures Analysis of Variance (ANOVA) was used to apply the multiple time series design. 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, Hispanic, Reduced Price Lunch, English Language Learner, and Special Education.

RESULTS

The evaluation of the Reading Camps was guided by a series of research questions that can be grouped into three components. The first component (Questions 1 and 2) pertained to the implementation and attitude. The second component gauged the program's impact on two measures: (a) students' alternative assessment scores (Question 3) and (b) on their likelihood of exceeding the cutoff on the alternative assessment required to be promoted (Question 4). The third component (Question 5) concerned the equitable distribution of growth among the various levels of selected demographic subgroups.

Return Rate

Principals at 35 schools were targeted to receive the Summer Implementation Survey, which was administered online during the summer session. Of those targeted, 82.6% (n=29) completed the survey. The return rate was therefore high enough to generalize the results to the target population.

Implementation

Implementation was gauged so that any observed effects on participants' achievement could be properly attributed to the program. An online survey entitled the Summer School Implementation Survey was used to make this determination, by measuring various aspects of the summer program including enrollment, planning, staffing, transportation, resources, materials, supplies, and specific curricular-factors (i.e.; dosage, frequency, and duration of treatment; engagement; and, perceived effectiveness). Several schools operated more than one program during the summer. Table 6 lists the number and percent of respondents that reported operating each program.

Table 6
Programs that Operated in Conjunction with the Reading Camps

Program	n	%
Algebra End of Course Remediation and Assessment	3	10.3
Extended School Year Services	2	6.9
Literacy for Rising Sixth Grade Students (Levels 1 and II on the FCAT 2.0, Grade 5 Reading Subtest)	3	10.3
Literacy for Rising Third Grade Students (below 50th percentile on the SAT-10, Grade 2 Reading Comprehension Subtest)	29	100.0
Web-Based Digital Learning for Third, Fourth, and Fifth Grade Students	28	96.6
Other (Middle Grades Course Remediation, Recapturing the Vision)		
Total	29	100.0

Note. Most schools reported operating three additional programs

All responding principals reported operating additional programs, 96.6% reported operating three or more, 24.1% reported operating four or more, and 3.5% reported operating five. The most prevalent program was Literacy for Rising Third Grade Students second grade tutoring followed by Web-Based Digital Learning for Third, Fourth, and Fifth Grade Students; and, Extended School Year services.

Literacy for Rising Sixth Grade Students and Algebra End of Course Remediation and Assessment each operated in 10.3% (n=3) of the schools. Two schools operated "other" programs (e.g., Middle Grades Course Remediation and Recapturing the Vision).

Despite the large number of programs that operated, the capacity of the schools was often reported to be below capacity. Nearly half the schools operated at or less than one-fifth of capacity, and three-quarters of the schools operated at less than half of capacity. The capacity of the school buses that transported the students was also addressed. Table 7 lists the number and percent of schools that reported capacity in each of six listed ranges. The table shows that the school buses were often better utilized than the schools with over one-third operating at near capacity. Nonetheless, the majority of the busses were reported to be less than half full.

Table 7
Categories of School Bus Capacity

	n	*%*
Less than 20% full	5	17.2
20 to 39% full	6	20.7
40 to 59% full	6	20.7
60 to 79% full	2	6.9
80% or more full	10	34.5
Total	29	100.0

The bulk of the implementation analysis was based on the results of 13 Likert type responses that gauged principals' perception of the adequacy of each of the five major dimensions of implementation: Planning, Resources, Transportation, Instructional Materials/Supplies, and Curriculum.

Response options within each dimension were classified as either positive or negative according to a rubric defined for each of the five implementation dimensions: Planning, Resources, Transportation, Instructional Materials/Supplies, and Curriculum. Table 8 lists the positive and negative response options for each implementation dimension.

Table 8
Response Option Categories Assigned to Each Implementation Dimension

	Response Option Category	
	Positive	Negative
Implementation Dimension	Response Option	Response Option
Planning	Adequately	Less than adequately
	More than adequately	
Resources	An appropriate level	Much less than needed
	More than needed	Somewhat less than need
	Much more than needed	
Transportation	Usually on time	Mostly not on time
	Mostly on time	Usually not on time
Instructional Materials/Supplies	An appropriate level	Much less than needed
	Much more than needed	
Curriculum	Average	Poor
	Good	Fair
	Very good	

Table 9 lists the number and percent of principals who responded positively and negatively to each item within each implementation dimension. A total line is also provided that summarizes the responses within each dimension. Total scores for each dimension are classified as: 0-49.9% (not implemented), 50-69.9% (partially implemented – shaded dark orange), and 70-100% (fully implemented – shaded light aqua).

Table 9
Principals' Responses to Selected Items Addressing Implementation

	Dimension/Item	Positive		Negative	
		n	%	*n*	%
	Planning				
5	How well was your summer location informed about the academic needs of the students who would be attending?	27	93.1	2	6.9
6	How well was your summer location informed of the number of students that would be attending?	28	96.6	1	3.4
7	How adequate were the instructional staffing levels at the school, relative to the number of students that attended	29	100.0	0	0.0
	Total	84	96.6	4	3.4
	Resources				
8	How would you characterize the amount of basic supplies (e.g., paper, ink, towels, etc.) available at the school?	11	37.9	18	62.1
9	How would you characterize the number of ancillary (e.g., clerical, cafeteria, janitorial, etc.) staff available at the school?	12	41.4	17	58.6
10	How would you characterize the number of security personnel available at the school?	21	72.4	8	27.6
	Total	44	50.6	24	49.4
	Transportation				
12	To what extent did the arrival of the buses typically align with the schedule of the school?	29	100.0	0	0.0
13	To what extent did the departure of the buses typically align with the schedule of the school?	27	93.1	2	6.9
	Total	56	96.6	2	3.4
	Instructional Materials/Supplies				
14	How sufficient was the amount of curricular materials (i.e., books, workbooks, manipulatives, etc.) available?	17	58.6	12	41.4
15	How sufficient was the amount of instructional supplies (i.e., computers, whiteboards, pencils, paper, etc.) available?	19	65.5	10	34.5
	Total	38	73.1	14	26.9
	Curriculum				
19	How would you rate the ability of the After the Bell to remediate different students with different learning problems in the same classroom at the same time?	16	66.7	8	33.3
20	How would you rate the engagement of students in the thematic subject matter covered by After the Bell?	24	100.0	0	0.0
21	How would you rate the effectiveness of After the Bell at addressing the skills deficits of the students?	21	87.5	3	12.5
	Total	61	84.7	11	15.3

Note. Total scores for each aspect are categorized as follows: 0-49.9% (not implemented), 50-69.9% (partially implemented – shaded dark orange), and 70-100% (fully implemented – shaded light aqua).

The table shows that nearly all the principals rated the planning, for the number and needs of students that would be attending their schools, positively. Staffing levels were assessed positively as a result. In terms of planning, the program may be considered to be fully implemented.

On the issue of resources, the principals were less sanguine with nearly around 40% reporting shortages of clerical staff and/or supplies. Nearly three-quarters of the principals did report having an adequate number of security staff. In terms of resources, the program may be considered to be partially implemented.

On the issue of transportation, nearly all principals reported that the arrival and departure of the school buses aligned with the school schedule. In terms of transportation, the program may be considered to be fully implemented.

On the issue of Instructional Materials/Supplies Approximately, 60% of the principals felt that the amount of curricular materials were adequate and two-thirds were satisfied with the amount of instructional supplies. In terms of instructional materials/supplies, the program may be considered to be partially implemented.

Curriculum was the primary area of inquiry and as such was addressed by four items one of which asked respondents to identify the program operating in their school. Of the principals queried, (82.7%, n=24) identified After the Bell as the program that was implemented, (13.8%, n=4) did not know or provided uninterpretable answers, and (3.4%, n=1) identified Voyager, the program that operated in prior years.

Of the principals who identified After the Bell, approximately two-thirds judged that it was able to remediate students with different learning problems within the same classroom. All felt that its subject matter was engaging to students and 87.5% indicated that it was effective at addressing the skill deficits of the students. On the issue of curriculum, the program may be considered to be fully implemented.

Finally, and perhaps most importantly, principals were asked to indicate the daily time allotted to After the Bell. Table 10 lists the number and percent of schools that reported operating After the Bell for each of the range of times provided.

Table 10
After the Bell Daily Dosage

	Schools	
Range (minutes)	*n*	%
151 or more	6	25.0
121 to 150	8	33.3
91 to 120	5	20.8
61 to 90	1	4.2
31 to 60	4	16.7
Total	24	100.0

Note. Recommended daily dosage is 150 minutes per day per program guidelines.

Although, all of the responding schools reported operating the program five days per week, more than 40% (10 of 24) reported operating the program for less than recommended time. Moreover, as noted previously, four additional schools reported not knowing the program that was operated,

and one school reported operating Voyager, the program that operated last year. Regarding, uniformity and nature of instructional delivery, the program does not appear to have been consistently implemented.

The attitudinal component of implementation measures principals' summative perceptions of the program. The issues addressed included technical support, effectiveness, satisfaction with the status quo, and suggestions for improvement. Table 11 lists the number and percent of principals that responded positively and negatively to the pertinent items.

Table 11
Principals' Attitudes Toward the Program

			Positive		Negative	
		N	*n*	*%*	*n*	*%*
30	How would you characterize the support provided by staff from curriculum and instruction?	29	24	82.8	5	17.2
31	How effective do you believe the Reading Camps were at helping students improve their reading skills?	29	29	100.0	0	0.0
32	To what extent do you agree that the program should continue to operate in its current form?	29	19	65.5	10	34.5

Note. Positive responses: (Average to Very Good and Agree to Strongly Agree). Negative responses: (Poor to Fair and Strongly disagree to Unsure).

The table shows that while three quarters of principals endorsed the level of technical support provided by District reading specialists, little more than a third were equally positive about the effectiveness of the program at improving the students' reading skills. Little more than half agreed or strongly agreed that the program should continue in its current form.

Space was also provided for principals to provide an open ended comment. Twenty-nine percent of the principals who completed the implementation items responded to this question as did twenty percent of those who did not. Three of the eight comments concerned time. One principal said,

> The program is precise and tailored to meet the specific needs of the learner. Additional time to implement the writing component and additional days for the students to work with the program prior to the summer assessment might be beneficial.
>
> Another principal expressed this need, "Additional instructional time is needed. Instructional time should be at least 3 hours. Lunch took at least a half hour." Still another principal voiced support for the program, noting that, "Teachers were able to see big improvements in students however, they need a lot more time."

Two of the eight comments concerned planning. One principal said,

> Hiring teachers was a huge challenge and this was very stressful for us and the Summer School leaders at the District level in coordinating with HR through the SAP requirements. Also, housing summer school at one location but having it technically processed at a separate

> location is challenging and given the timeline we had to prepare this time for registration, transportation, etc. we did the best we could.

Another principal explained,

> Parents had already made summer plans for their students and this wonderful opportunity was missed by many students. The After the Bell program was great for the students but materials were extremely limited.

Two of the comments concerned resources, one principal said,

> There needs to be a budget for supplies, as well as an hourly custodian. The Summer Reading Camp also should be funded at a lower student to teacher ratio, preferably 1:10, because the students who attend are very deficient, and they need intense intervention. The typical classroom student to teacher ratio is not effective for these deficient children and should be lowered for the summer session to truly provide the students with remediation.

Another principal felt that the needs of special student populations were not adequately addressed, "Provide materials with a similar format to the instruction provided throughout the school year. Students that are in SPED and/or ELL need a modified curriculum to support their needs."

In sum, while principals generally gave positive reports of the core aspects of implementation (i.e., planning, resources, transportation, instructional materials/supplies and curriculum). However, the availability resources and instructional materials/supplies were rated as sufficiently wanting to cause them to be judged as only partially implemented. These sentiments were echoed in open ended comments in which lack of time constituted an additional cause of concern.

Finally, although, the Reading Camps curricula were rated as effective, over one-third of respondents did not agree that they should continue in their present form. Notwithstanding the foregoing, an analysis of program dosage revealed such wide variation that almost no two schools operated the program the same way.

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 demographic characteristics and then compared the groups' adjusted posttest scores, while controlling for students' initial ability as measured by their adjusted pretest scores. A detailed description of this process may be found in Appendix B.

Status

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.

Table 12 displays the results of this process and lists for each predictor; descriptive statistics (i.e., mean and standard error in original scale score units) of its unstandardized regression weight, the standardized weight, and the results of a *t*-test that gauges the statistical significance of the weight.

Table 12
Regression Analysis of the Adjusted Scores

	Unstandardized Coefficients			
Effects	*M*	*SE*	Standardized Coefficients	*t*
Intercept	584.30	2.56		229.53***
Pretest	1.07	0.05	.47	20.53***
Program	-0.37	2.61	-.00	-0.14
Program x Pretest	--	--	--	--

Note. The weights represent the influence on the criterion variable of a unit change in the predictor. All predictors are expressed as deviations from the sample mean (i.e., grand-mean centered). 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. The pre- and posttest scores are adjusted for demographics.
*$p < .05$. ***$p < .001$

The table shows there a non-significant 0.37 point difference between the mean posttest scores of participants and non-participants, 583.93 and 584.30, respectively. The effect size of this difference (d = -.02) has weak practical significance[2] and the non-significance of the interaction term indicates that program effect did not vary with the pretest.

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' posttest scores after their pretest scores were taken into account. The posttest scores were converted to pass/fail outcomes based on whether or not the scores met or exceeded the criterion for promotion (i.e., 45th percentile).

Table 13 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). Demographic characteristics were also included because cutoff scores could not be adjusted to account for them.

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

Table 13
Logistic Regression Analysis of the Posttest Scores

Effect	*B*	S.E.	Odds Ratio	Wald
Intercept	-2.05	0.37	0.13	30.20***
Program	0.39	0.37	0.66	1.27
Pretest	0.09	0.01	1.10	91.53***
Program x Pretest	--	--	--	--
Disabled	-1.00	0.28	0.37	12.47***

Note. Demographic predictors are not displayed. 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.
*** $p < .001$.

Of the 1,499 students in this study who took the AATGP, only 9.6% (n=136) of the students who attended the Reading Camps and 13.3% (n = 10) of the students in the comparison group passed the test. The odds of passing the test, for students who completed the Reading Camps relative to those who did not participate (0.66 to 1) were not statistically significant, once their pretest scores and demographic characteristics were taken into account. Thus, completing the Reading Camps did not improve students' chances of passing the AATGP.

To summarize, students who completed the program scored 0.37 scale score points lower than students who did not participate in the program, when their pretest scores and demographic characteristics were taken into account. Moreover, students who completed the program had 1.51 lesser odds (1 ÷ 0.66) of being promoted than students who did not participate, once their pretest scores and demographic characteristics were taken into account.

Student Growth

This section of the evaluation examined participating students' growth on the alternative assessment between the spring and summer administrations, overall and among the levels of selected demographic subgroups. Table 14 list for each subgroup of the pre- and posttest, the sample size, and the mean and standard deviation test.

Table 14 shows the presence of pre- to post- test gains at each level of each subgroup. Subgroup pretest means ranged from a low of 544.3 (SPED) to a high of 560.6 (non-FRL) and averaged 551.6. Subgroup posttest means ranged from a low of 571.3 (SPED) to a high of 593.2 (non-FRL) and averaged 583.9. Pre- to posttest growth estimates ranged from a low of -6.69 (SPED) to a high of 34.5 (Female) and averaged 32.3.

Table 14
Alternative Assessment Scores of Students who Completed the Reading Camps by Demographic Subgroup

		Pretest [a]		Posttest		Estimated Growth	
Group	*N*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Race							
Black	595	549.4	27.34	582.5	25.57	33.16	28.83
Non-Black	829	553.1	28.67	584.7	26.57	31.63	26.22
Difference [b]		-3.73	28.12	-2.20	26.16	1.53	27.34
English Language							
ELL	724	551.5	28.35	582.3	25.82	30.86	26.04
Non-ELL	700	551.6	28.00	585.4	26.45	33.73	28.57
Difference [b]		-0.19	28.18	-3.06	26.13	-2.87	27.31
Gender							
Female	549	553.7	28.83	588.2	25.92	34.48	27.07
Male	875	550.2	27.68	581.1	25.97	30.88	27.43
Difference [b]		3.49	28.13	7.08	25.95	3.59	27.29
Reduced Price Lunch							
FRL	1,347	551.0	27.87	583.3	26.12	32.25	27.34
non-FRL	77	560.6	31.87	593.2	25.39	32.60	27.59
Difference [b]		-9.56	28.10	-9.91	26.08	-0.35	27.35
Special Education							
SPED	307	544.3	27.61	571.3	25.83	27.02	27.01
non-SPED	1,117	553.5	28.01	587.2	25.21	33.71	27.26
Difference [b]		-9.22	27.92	-	25.35	-6.69	27.34
Overall	**1,424**	**551.6**	**28.17**	**583.9**	**26.17**	**32.3**	**27.34**

Note. Standard deviations of differences are pooled values.
[a] Pretest scores are corrected for difficulty and transformed to the scale of the posttest.
[b] First listed level minus the second listed level

Although, overall growth was seen, differences in growth between the levels of various subgroups were evident. Moreover, achievement gaps between the levels of the subgroups were present on both the pre- and the post- test. Although, many differences were identified, the statistics displayed above do not give the statistical or practical import of those differences. Therefore, a repeated measures ANOVA was conducted to accomplish both those goals and present the results concisely.

Table 15 summarizes these results and lists for each source of variation, the F statistic used to gauge its statistical significance, its practical significance (effect size), $\eta_{p.}$[3] Significant effects, identified by examining the F statistic, are indicated by one or more asterisks.

The between-subjects effects describe differences between the average of the pre- and posttest scores of students at different levels (e.g., ELL vs. non-ELL) of a subgroup. Significant effects indicate achievement gaps among students' status scores.

The within-subjects effect of Time describes growth from pre- to post- test. The within-subjects interaction terms (e.g., Time x Female) describes differences between the pre- to post- test growth of students at different levels (e.g., Male vs. Female) of a subgroup. Significant effects indicate the presence of achievement gaps among students' growth scores.

[3] Cohen (1988) has classified the practical significance of the effect size η as .10 (weak), .24 (moderate), and .37 (strong).

Table 15
Repeated Measures Analysis of Variance:
Effect of Selected Demographic Factors on Students' Alternative Assessment Scores

Source	*df*	*F*	η_p
	(Between Group)		
Black	1	16.77***	.11
English Language Learner (ELL)	1	10.40***	.09
Female	1	8.86**	.08
Reduced Price Lunch (FRL)	1	12.66***	.09
Special Education (SPED)	1	76.19***	.23
S within-group error	1,418	(1,021.50)	
	(Within Group)		
Time	1	327.63***	.43
Time x Black	1	0.14	.01
Time x ELL	1	3.05	.05
Time x Female	1	3.64	.05
Time x FRL	1	0.07	.01
Time x SPED	1	12.35***	.09
Time x S within-group error	1,418	(369.27)	

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.

Table 15 shows that significant difference in students' between-group (i.e., average) scores were identified for each of the subgroups. All of the η effect sizes represent weak levels of practical significance. There was also a significant strong effect for time, indicating that that substantial student growth from pre- to posttest was experienced. Though caution must be exercised in quantifying this growth in because the scores were not formally equated, the effect size[4] of participating students' gains (d=1.18) is nearly as large as the retrofitted gains of third grade retainees (d=1.38) on the two most recent administrations of the FCAT 2.0 (Urdegar, 2012).

The table also shows significant within-subject interactions for the SPED subgroup. This indicates there were significant differences in the pre- to post test growth of students at different levels of that subgroup.

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 on all the subgroups that mirrored the results of other studies (e.g., Casanova, Garcia-Linares, de la Torre, & de la Villa Carpio, 2005); Eamon, 2005; Urdegar, 2008). Strong growth from pre- to post- test was seen for participating students as a whole. However, significant differences in the pre- to post test growth of students at different levels of the SPED subgroup, indicate that program's benefits were not equitable distributed for those students.

[4] Based on Cohen's *d* as derived from a paired sample *t* test converted to the independent case (Cortina & Nouri, 2000, p. 50).

DISCUSSION

The district undertook an evaluation to gauge whether students who completed the Reading Camps 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.

Summary

The evaluation of the Reading Camps described the operation and assessed multiple measures of outcome for the program that operated during its tenth 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. **Were the Reading Camps implemented as planned?**

   While principals gave positive accounts of planning, transportation, and curriculum/instruction, and rated all three parts of the curricululum as effective; the availability of materials and supplies was rated as somewhat wanting. An analysis of *After the Bell* program dosage revealed wide variation so that almost no two schools seem to have operated the program the same way. Although most principals endorsed the level of technical support provided by District reading specialists, they were less enthusiastic about the effectiveness of the program at improving the students' reading skills. As such, the program could not be judged to be consistently implemented, and the following results must be reviewed with this in mind.

2. **Did students who participated in the Reading Camps score higher on the summer alternative assessment than students who did not?**

   The findings demonstrate that students who completed the program did not score 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 practical significance and was not statistically significant. Therefore, the program did not provide an advantage to students who participated, over those who did not attend, but only registered to take the test.

3. **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?**

   The results show that students who completed the program were not significantly more likely to meet the cutoff for promotion than students who did not participate in the program, when students' pretest scores and demographic characteristics were taken into account.

4. **Did students who attended the Reading Camps show growth on the alternative assessment from spring to summer?**

The results show significant strong overall growth from the pretest to the posttest. However, the growth seen by Special Education (SPED) students growth was significantly less than was seen by their non-SPED counterparts.

Conclusions

While participants' reading scores and chances of passing the alternative assessment test required for promotion were not significantly better than those of non participants, significant growth between the pretest and the posttest was seen. Large variations in the implementation of the program may have lessened its effectiveness.

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.

Casanova, P. F., Garcia-Linares, M. C., de la Torre, M. J., & de la Villa Carpio, M. (2005). Influence of family and sociodemographic variables on students with low academic achievement [Electronic version]. Educational Psychology, 25(4), 423-435.

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

Department of Language Arts/Reading (2012). *Summer Waves of Leaning program description of the remedial components for third grade students.* (Email transmittal). Miami, FL: Miami-Dade County Public Schools.

Department of Summer Services (2012). *Implementation Documentation: 2012 Summer Waves of Learning* [Electronic version]. Miami, FL: Miami-Dade County Public Schools. Retrieved July 31, 2012, from http://summerschool.dadeschools.net/pdfs12/SID12.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. (2008). *Critical information for parents of third graders* [Electronic version] Retrieved August 13, 2012 from http://flbsi.org/pdf/2008ReadtoLearn.pdf

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 (2012). Retrieved August 13, 2012, from http://www.flsenate.gov/Laws/Statutes/2012/1008.25

Urdegar, S.M. (2008). *Beyond Fidelity: Relating educational practices and their determinants to student learning gains.* (Doctoral dissertation, Florida Atlantic University, 2008).

Urdegar, S.M. (2012). *A within group analysis of retained third graders' performance on the FCAT-SSS*. Unpublished data analysis. Miami, FL: Miami-Dade County Public Schools.

Appendix A

Summer School Implementation Survey

Summer Reading Camps Implementation Survey

Instructions

This survey is designed to assess various aspects of the programs that operate at the school to which you are assigned during the summer. The information that you provide will be used to furnish formative feedback and inform future planning in the areas of logistics, human resources, and curriculum. Your impressions are very important to the future success of this initiative. If you have any questions or need assistance with this survey you may contact Dr. Steven M. Urdegar at (305) 995-7538. Please reflect honestly on your experience as an administrator and answer each item in the survey that follows to the best of your ability. Thank you in advance for your cooperation.

Introduction

The items that follow pertain to the school to which you are assigned during the summer.

1. Which of the following choices best describes your role at the school?

(Select only one.)

- ☐ Principal
- ☐ Assistant Principal
- ☐ Other (specify):

2. Approximately, how many students are enrolled in this school during the regular school year?

(Provide one response only.)

3. Approximately, how many students participate in one or more of the summer programs?

(Provide one response only.)

Planning

The items that follow pertain to ALL the programs operating at the school during the summer session.

4. Which programs operate at the school?

(Select all that apply.)

- ☐ Third Grade Summer Reading Camps
- ☐ Literacy for Rising Third Grade Students (below 50th percentile on the SAT-10, Grade 2 Reading Comprehension Subtest)
- ☐ Literacy for Rising Sixth Grade Students (Levels 1 and II on the FCAT 2.0, Grade 5 Reading Subtest)
- ☐ Web-Based Digital Learning for Third, Fourth, and Fifth Grade Students
- ☐ Algebra End of Course Remediation and Assessment
- ☐ Credit Recovery
- ☐ Extended School Year Services
- ☐ Other (specify):

5. How well has your present summer location been informed about the academic needs of the students who would be attending?

(Select only one.)

- ☐ Less than adequately
- ☐ Adequately
- ☐ More than adequately
- ☐ Not applicable

6. How well has your summer location been informed of the number of students that would be attending?

(Select only one.)

- ☐ Less than adequately
- ☐ Adequately
- ☐ More than adequately
- ☐ Not applicable

7. How adequate are the instructional staffing levels at the school, relative to the number of students that attended

(Select only one.)

- ☐ More than 20% understaffed
- ☐ 10 to 19% understaffed
- ☐ Appropriately staffed
- ☐ 10 to 19% overstaffed
- ☐ More than 20% overstaffed

Resources

The questions that follow pertain to all the programs operating at the summer school.

8. How would you characterize the amount of basic supplies (e.g., paper, ink, towels, etc.) available at the school?

(Select only one.)

- ☐ Much less than needed
- ☐ Somewhat less than needed
- ☐ An appropriate level
- ☐ More than needed
- ☐ Much more than needed

9. How would you characterize the number of ancillary (e.g., clerical, cafeteria, janitorial, etc.) staff available at the school?

(Select only one.)

- ☐ Many fewer than needed
- ☐ Somewhat fewer than needed
- ☐ An appropriate level
- ☐ More than needed
- ☐ Many more than needed

10. How would you characterize the number of security personnel available at the school?

(Select only one.)

- ☐ Many fewer than needed
- ☐ Somewhat fewer than needed
- ☐ An appropriate level
- ☐ More than needed
- ☐ Many more than needed

Transportation

11. On average, how full are the school buses used to transport the students to and from the school?

(Select only one.)

- ☐ Less than 20% full
- ☐ 20 to 39% full
- ☐ 40 to 59% full
- ☐ 60 to 79% full
- ☐ 80% or more full

12. To what extent does the arrival of the buses typically align with the schedule of the school?

(Select only one.)

- ☐ Mostly not on time
- ☐ Usually not on time
- ☐ Usually on time
- ☐ Mostly on time

13. To what extent does the departure of the buses typically align with the schedule of the school?

(Select only one.)

- ☐ Mostly not on time
- ☐ Usually not on time
- ☐ Usually on time
- ☐ Mostly on time

Instruction

The items that follow pertain ONLY to the Third Grade Summer Reading Camps

14. How sufficient is the amount of curricular materials (i.e., books, workbooks, manipulatives, etc.) available?

(Select only one.)

- ☐ Much less than needed
- ☐ Somewhat less than needed
- ☐ An appropriate level
- ☐ More than needed
- ☐ Much more than needed

15. How sufficient is the amount of instructional supplies (i.e., computers, whiteboards, pencils, paper, etc.) available?

(Select only one.)

- ☐ Much less than needed
- ☐ Somewhat less than needed
- ☐ An appropriate level
- ☐ More than needed
- ☐ Much more than needed

Branch1

16. Does After the Bell (developed by Scholastic) operate in your school?

(Select only one.)

- ☐ Yes
- ☐ No
- ☐ Don't know

Curriculum

17. How many days per week is instruction in After the Bell typically delivered to students?

(Select only one.)

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

18. On the days in which it is used, for how many minutes daily was After the Bell provided?

(Select only one.)

- ☐ 30 or less
- ☐ 31 to 60
- ☐ 61 to 90
- ☐ 91 to 120
- ☐ 121 to 150
- ☐ 151 or more

19. How would you rate the ability of After the Bell to remediate different students with different learning problems in the same classroom at the same time?

(Select only one.)

- ☐ Poor
- ☐ Fair
- ☐ Average
- ☐ Good
- ☐ Very Good

20. How would you rate the engagement of students in the subject matter covered by After the Bell?

(Select only one.)

- ☐ Poor
- ☐ Fair
- ☐ Average
- ☐ Good
- ☐ Very good

21. How would you rate the effectiveness of After the Bell at addressing the skills deficits of the students?

(Select only one.)

- ☐ Poor
- ☐ Fair
- ☐ Average
- ☐ Good
- ☐ Very Good

Branch4

22. Did any other component of the Reading Camps other than After the Bell operate in the school during the regular school day of the present summer session?

(Select only one.)

- ☐ Yes
- ☐ No
- ☐ Don't know

Other

23. Please provide the name and a description of the program that includes publisher/developer, series, and materials used.

(Provide one response only.)

24. Which of the following skills were addressed by the program?

(Select all that apply.)

- ☐ Phonemic Awareness
- ☐ Phonics
- ☐ Vocabulary development
- ☐ Fluency
- ☐ Reading comprehension
- ☐ Writing
- ☐ Test preparation
- ☐ Other (specify):

25. How many days per week was instruction in the program typically delivered to students?

(Select only one.)

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5

26. On the days in which it was used, for how many minutes daily was the program provided?

(Select only one.)

- ☐ 30 or less
- ☐ 31 to 60
- ☐ 61 to 90
- ☐ 91 to 120
- ☐ 121 to 150
- ☐ 151 or more

27. How would you rate the ability of the program to remediate different groups of students with different learning problems in the same classroom at the same time?

(Select only one.)

- ☐ Poor
- ☐ Fair
- ☐ Average
- ☐ Good
- ☐ Very Good

28. How would you rate the engagement of students by the program?

(Select only one.)

- ☐ Poor
- ☐ Fair
- ☐ Average
- ☐ Good
- ☐ Very good

29. How would you rate the effectiveness of the program at addressing the skills deficits of the students?

(Select only one.)

- ☐ Poor
- ☐ Fair
- ☐ Average
- ☐ Good
- ☐ Very Good

Summary1

The question that follow pertain to your overall impression of all aspects of the Third Grade Summer Reading Camps

30. How would you characterize the technical support provided by staff from curriculum and instruction?

(Select only one.)

- ☐ Poor
- ☐ Fair
- ☐ Average
- ☐ Good
- ☐ Very Good
- ☐ Not applicable

Summary2

31. How effective do you believe the Third Grade Summer Reading Camps were at helping students to improve their reading skills?

(Select only one.)

- ☐ Poor
- ☐ Fair
- ☐ Average
- ☐ Good
- ☐ Very Good
- ☐ Not applicable

32. To what extent do you agree that the program should continue to operate in its current form?

(Select only one.)

- ☐ Strongly disagree
- ☐ Disagree
- ☐ Unsure
- ☐ Agree
- ☐ Strongly agree

33. Optionally, use the space below to provide any suggestions you might have on how the Summer Reading Camps can be improved.

(Provide one response only.)

Appendix B

Statistical Addendum

Demographic Adjustment

Regression analysis[5] to separately estimate the effect of selected demographic characteristics (i.e., gender, ethnicity, reduced price lunch eligibility, ELL status, and SPED classification) on each student's pretest and posttest scale score in order to account for the influence of those factors when contrasting the groups' performance. 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.

The results of this process are displayed in Table B1. The table lists the adjustment 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). Standardized weights, which are scaled to zero mean and unit variance, are independent of the scale of the test.

Table B1
Regression Analysis used to Demographically Adjust the Students' Test Scores

	Pretest				Posttest			
	Unstandardized (B)				Unstandardized (B)			
Predictor	M	SE	Standardized (β)	t	M	SE	Standardized (β)	t
Intercept	166.81	1.28		130.11 ***	601.30	2.99		201.29 ***
Black	-2.60	0.72	-0.12	-3.62 ***	-7.29	1.65	-0.14	-4.42 ***
English Language Learner	-1.39	0.70	-0.06	-1.98 *	-7.38	1.60	-0.14	-4.60 ***
Female	--	--	--	--	5.10	1.34	0.09	3.81 ***
Reduced Price Lunch	-3.83	1.24	-0.08	-3.08 **	-9.56	2.85	-0.08	-3.36 ***
Special Education	-3.72	0.69	-0.14	-5.41 ***	-16.12	1.58	-0.25	-10.17 ***

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. Statistics for predictors that were not entered into the stepwise model are shown as dashes

* p < .05. **p < .01. *** p < .001.

Table B1 shows that the Black, English Language Learner, and Special Education classifications each have significant influence on the pre- and post- tests. Classification as Female has a significant influence on the posttest beyond those previously listed.

On the pretest, Black students scored 2.60 points lower than students who were not Black, students who were English Language Learners (ELL) scored 1.39 points lower than students who were not ELL, students who were eligible for Reduced Price Lunch score 3.72 points lower than students who were not eligible,. and students with Special Education (SPED) classifications scored 3.72 points lower than those who were not so classified.

[5] 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).

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, gender ethnicity, and English language learner status have significantly different influences on the pretest and the posttest.

The standardized weights allows the impact of each predictor to be compared directly, even though the scale of the pre- and post- test are different, by expressing the impact of the predictors on the outcome, when each is scaled to zero mean and unit variance. This causes the intercept to be fixed at zero. All the predictors except ELL and SPED have similar influences on the pre- and post- test with Black being the strongest followed by Reduced Price Lunch. However, the influence of the ELL and SPED classifications on the posttest are nearly twice as large as their influence on the pretest.

As such, any imbalance between the groups on those characteristics could potentially mask the treatment effect. Therefore, an estimate was produced for each student's pre- and posttest score and used to adjust their actual scores with scores that represented the pretest and posttest scores that s/he would have had had s/he had the demographic characteristics of the sample.

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.

A difficulty correction was applied to equate the probabilities that the norm samples for the pretest and posttest would meet/exceed the passing criteria defined by the state for the Reading Camps. The difficulty correction was determined by converting the probability of passing the pretest into an odds ratio and taking its logarithm. The difficulty correction (i.e., -0.201) 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.