Striving Readers Year 5 Project Evaluation Report: Ohio
An Addendum to the Year 4 Report

Prepared by[1]

William E. Loadman, Ph.D.
Raeal J. Moore, Ph.D.
Weijia Ren, M.A.
Jing Zhu, Ph.D.
Jing Zhao, M.A.
Richard Lomax Ph.D.

The Ohio State University

Submitted to the Office of Elementary and Secondary Education

December, 2011

[1] Special thanks to Charles Okonkwo for his classroom observation work. Special thanks to Hua Xu for his work on selected analyses.

Table of Contents Page

List of Tables Page

List of Figures Page

I. Executive Summary of Findings: Targeted intervention Implementation and Impact

This report is an addendum to Ohio's Year 4 Striving Readers program evaluation report and contains data gathered and analyzed in year 5 of the project. The narrative here focuses on an update to the targeted intervention – both program implementation and impact, with the whole school evaluation omitted. The Year 4 report discusses more contextual information (e.g., measures used, psychometric analysis of the primary outcome); this report can be found at the following location: http://www2.ed.gov/programs/strivingreaders/performance.html.

A. Implementation

From October 2006 to June 2011, the Ohio Department of Youth Services (ODYS) implemented Scholastic's Read 180 program in the seven DYS high schools. Read 180, a daily 90-minute structured reading program, is composed of five components – whole group, individualized learning, computer activities, small group, and wrap up. The program was offered to students randomly assigned to treatment conditions; these students were then assigned to the appropriate high school based on levels of offense. To be eligible, a student had to have a score below grade level (approximately 1000 Lexile points), but above "below basic" level (a Lexile score of 200 or less) at baseline on the Scholastic Reading Inventory (SRI). In the five years of program implementation, this resulted in 1982 youth (1058 Read 180 assigned, 924 traditional English assigned) housed at ODYS, which is a part of the targeted intervention portion of this evaluation.

To assess the fidelity of the Read 180 implementation, professional development attendance records, number of minutes in Read 180 instruction, evaluation team observational records, and Scholastic in-class assessments and feedback were collected. PD attendance and number of minutes in Read 180 instruction were the sole data sources for evaluating the fidelity of the professional development and instructional models, respectively.

Program implementation for instruction in Year 5 varied across facility. Two facilities (40%) were rated as either "moderate" or "high" in instructional implementation; the remaining three facilities were rated as "needing improvement". It was a challenge for teachers in each facility to execute the entire 90 minutes, a pattern consistent across five years of program implementation. However, teachers allocated more time for Read 180 instruction in the first two years of the project, in general, relative to the last three years of the project. On the other hand, all facilities (n=5)[2] were rated as either "high" or "moderate" in PD attendance implementation. There was limited consistency within a facility on these two implementation indicators.

B. Impact

The Read 180 program had an impact on struggling readers based on two outcome measures, the SRI and the CAT. A series of Intent-To-Treat (ITT) analyses – both cross sectional and longitudinal - were conducted to determine whether Read 180 improved the reading performance for youth reading below grade level. Using SRI as an outcome, youth who were supposed to receive two or more quarters of Read 180 instruction (n=677) outperformed, on average, youth in the traditional English classes (n=568) by an additional average gain of approximately 59 Lexile points after two quarters of intended treatment, based on the cross sectional analysis. Additional analyses, including

[2] In year 5, the number of possible facilities was reduced from seven in the first two years of the project to five facilities in year 5.

a longitudinal ITT model, found similar gains. Youth in both the Read 180 and traditional English classes still however remained below their reading grade level even after exposure to either of the English curricula. The gain across two quarters of treatment represents, however, more than the projected gain after one year of treatment according to the age/grade youth specified by Scholastic.

Since the SRI test is more often practiced by the youth in Read 180 classes and its psychometric properties are not well-established for the targeted population, using the California Achievement Test (Read CAT) as a second outcome measure was employed. Two ITT, HLM models were estimated – an outcome measure that was obtained after a year of being housed at ODYS and an alternative outcome measure that was the last assessment. Using a cross-sectional ITT, HLM model those assigned to the Read 180 intervention did improve significantly more in reading ability relative to their English assigned counterparts when the Read CAT assessment after a year of treatment was utilized. No statistically significant impacts were found when the last Read CAT assessment was utilized.

II. Evaluation of the Implementation of the Targeted Intervention: Year 5

A. Summary of the design

ODYS's targeted intervention implementation study centers on four over-arching evaluation questions:

(1) What was the level of implementation and facility level variability of professional development/support for coaches, Read 180 teachers/Aids, and principals in Years 1 through 5?
(2) What was the level of implementation and facility level variability of classroom instruction in Years 1 through 5?
(3) How did the level of implementation and variability of professional development/support for coaches, Read 180 teachers/Aids, and principals differ across Years 1 through 5?
(4) How did the level of implementation and variability of classroom instruction differ across Years 1 through 5?

The first evaluation question is answered using Professional Development (PD) attendance records provided by ODYS. The second evaluation question is addressed by: (a) the teacher logs recording daily time allocations per class, (b) weekly observations by the project evaluators, and (c) quarterly visits by a representative from Scholastic who visits each of the seven high schools to provide technical assistance to the instructional staff and observe the quality of program implementation. The classroom observations objectives conducted by project evaluators varied across years. The intent of these observations will be detailed below. The third and fourth questions are answered by comparing these collected data across the five years of program implementation.

B. Summary of the results

In year 5, one Read 180 professional development activity (4 hours) was available for the Read 180 teachers, aides, and literacy coaches. This session was an interactive professional development session with the Ohio State University evaluation team. Here the evaluation team presented the targeted intervention Year 4 results at the aggregated level.

Since there was a teacher, aide and literacy coach for each facility in Year 5, Table 1 presents whether that individual attended the one available Read 180 session (100%) or not (0%). Teachers, aides, and literacy coaches attending the Read 180 session were consistent across facilities. All of the facilities had a high percentage of attendance across the three Read 180 staff except facility 5, where the teacher did not attend the PD session.

The total percentage is the average percentage of attendance aggregated across the three Read 180 personnel in each facility. Most facilities had a 100% attendance across all three Read 180 personnel, resulting in a "high" level of implementation. Facility 5 had a 66.67% attendance and was rated as "moderate" in professional development attendance implementation using the scale defined below.

High = 75% - 100%
Moderate = 50% - 74%
Needs Improvement = < 50%

Table 1.Targeted Intervention Read 180 Professional Development Activities Attendance by Facility Year 5

Facility	% Teacher	% Aide	% Literacy Coach	% Total	Level
2	100.00	100.00	100.00	100.00	High
4	100.00	100.00	100.00	100.00	High
5	0.00	100.00	100.00	66.67	Moderate
7	100.00	100.00	100.00	100.00	High
8	100.00	100.00	100.00	100.00	High
Total	80.00	100.00	100.00	93.33	High

The amount of Read 180 instruction for each facility disaggregated by quarter in Year 5 was summarized in Table 2. As the project ended in the end of spring 2011, only three quarters of instruction were recorded. Although each facility had between one and three sections of Read 180 (taught by the same teacher and aide) this table aggregates instruction across these sections. Each facility's implementation of instruction appeared to vary and had different patterns across quarters.

Table 2. Average Minutes of Instruction Aggregated Across Blocks by Quarter and Facility in Year 5.

	Fall 2010	Winter 2011	Spring 2011	Average	Level
2	54	55	49	53	Needs Improvement
4	63	64	60	62	Needs Improvement
5	82	67	75	75	Moderate
7	68	55	67	63	Needs Improvement
8	79	82	--[a]	81	High
Total	69	65	63	67	Needs Improvement

[a] Data are unavailable for facility 8 for spring 2011.

Note: Facility 1 and facility 3 were closed at the end of project Years 3 and 4 respectively.

It was difficult for the 5 facilities to meet the 90-minute instruction model, which is also a problem evidenced in the prior three project years. Facilities 2, 4, and 7 had the least amount of average reported instruction in Year 5 with 53, 62 and 63 minutes respectively; these facilities were rated as "needs improvement" in instructional implementation. Facility 5 was rated as "moderate" in instructional implementation with 75 average minutes of instruction. A few instances occurred where students were not in school (e.g., fire drills, weather calamity or facility-wide lock downs). Teachers more frequently utilized Read 180 instruction time to complete other building wide objectives. For example, the Read 180 classes, because of the computer access, were used to test students (i.e., OGT, SRI and Rskills). Students sometimes watch movies, attend assemblies or had "fun" days to replace Read 180 instruction. Finally, in most situations, if a teacher was absent, students were either directed to the library or monitored and instructed using non-Read 180 material by a substitute teacher. Only facility 8 was rated as "high" in instructional implementation with an average of 82 minutes a day.

Implemented instructional time aggregated across quarters was rated for each facility utilizing the following rubric:

High = 80 and more minutes of instruction

Moderate = 74-79 minutes

Needs improvement = 73 and less minutes of instruction

In Year 5 one classroom observer observed at least two schools a week. Table 3 presents the frequency of observations by facility. The number of observations conducted by facility correlates with the number of Read 180 classes offered in a given day. Facility 2 had the most classes offered (e.g., three classes) while Facility 8 had the fewest number of offered Read 180 courses (e.g., one class). There were a total of 54 Read 180 classroom observations completed in Year 5.

Table 3 Read 180 classrooms observed by facility: Year 5

Facility	Frequency	Percent
2	18	33.3
4	12	22.2
5	8	14.8
7	9	16.7
8	7	13.0
Total	54	100

Table 4. Average number of minutes of observed total instruction by facility: Year 5

Facility	N	Mean	S.D.
2	18	91.9	7.3
4	12	90.8	12.7
5	8	92.1	22.6
7	9	90.2	10.9
8	7	83.7	3.4
Total Instruction	54	90.3	12.0

Table 4 illustrates that overall Read 180 teachers and aides are implementing 90 minutes of total Read 180 instruction on the days of observation. This amount of instruction is pretty consistent across four of the five facilities with one facility (Facility 8) lagging slightly behind (M = 83.7). Noteworthy is the variability of program minutes, particularly with Facility 5. In general, according to these self-reported measures, facilities are implementing the total number of Read 180 instruction minutes as Scholastic has specified. We also wanted to triangulate the start time data collected by Read 180 staff (see Table 2) with the total number of minutes reported by the classroom observers. Read 180 staff reported implementing significantly fewer average Read 180 instructional minutes in Facility 2 (M = 53 minutes) relative to the average number of minutes

obtained from classroom observations (M = 91.9minutes). This pattern was present too with Facilities 4, 5, and 7. Facilities show consistent results when comparing the two data sources.

To unpack the total number of Read 180 minutes, Table 5 presents the percentage of classes where a given rotation was observed to be implemented. It appears that overall and for each facility teachers are frequently omitting wrap up. Only 22.2% of the classes observed implemented this rotation; Facilities 4 and 8 on the days observed omitted it entirely. Teachers appeared to implement the remaining rotations consistently, accounting for the 90 minutes of implemented instructions shown in Table 4.

Table 5. Frequency of Read 180 classes observed implementing each rotation: Year 5

	Facility											
	2 (*n*=18)		4 (*n*=12)		5 (*n*=8)		7 (*n*=9)		8 (*n*=7)		Total (*n*=54)	
Rotation	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
WG	12	100.0	12	100.0	8	100.0	9	100.0	7	100.0	54	100.0
SG	15	83.3	12	100.0	8	100.0	7	77.8	7	100.0	49	90.7
CR	12	100.0	12	100.0	7	87.5	9	100.0	7	100.0	53	98.1
IR	17	94.4	12	100.0	7	87.5	9	100.0	7	100.0	52	96.3
WU	3	16.7	0	0.0	3	37.5	6	85.7	0	0.0	12	22.2

WG = Whole Group; SG = Small Group; CR = Computer Rotation; IR = Individual Reading: WU = Wrap Up

C. Year 1 – Year 5 implementation

Changes in the level of implementation from Year 1 to Year 5. Teacher, aide, and principal professional development attendance across the first two years was relatively consistent, with a "high" level of implementation reported. However, in Year 3 the level of professional development attendance showed more facility level variability and challenges in implementation for some facilities. In Years 4 and 5, consistencies across facilities emerged. In terms of the amount of implemented instruction, no facilities maintained consistencies across the five years. Facilities 2 and 5 appeared to struggle in instructional implementation with consecutive "needs improvement" ratings while Facility 8 was frequently rated as "high". Table 6 summarizes the five years of program implementation.

Implications for impact analysis. Variation in program implementation across the sites and across years may have consequences for the impact analyses. Specifically, youth who were only exposed to the Read 180 program in the third year might be negatively influenced by limited minutes allocated to each of the five components, particularly in Facilities 2 and 5. It should however be noted that the high student mobility across facilities makes it a challenge to determine the degree to which students would be influenced by program implementation variations. Overall, implementation of the Read 180 targeted intervention generally occurred at a moderate level as judged by the external evaluators, notwithstanding an aberration at a given facility.

Table 6. Summarized Ratings of Targeted Intervention Professional Development and Instruction

	Professional Development					Instruction				
Fac.	Year 1	Year 2	Year 3	Year 4	Year 5	Year 1	Year 2	Year 3	Year 4	Year 5
1	M	H	NI	N/A	N/A	NI	M	M	N/A	N/A
2	M	H	H	H	H	M	NI	NI	NI	NI
3	H	H	M	N/A	N/A	M	H	H	M	N/A
4	H	H	H	M	H	H	H	M	M	NI
5	H	H	NI	H	M	M	M	NI	NI	M
7	H	H	M	H	H	M	M	H	M	NI
8	H	H	NI	H	H	M	H	H	H	H
Total	H	H	M	H	H	M	H	M	M	NI

Note: NI= needs improvement; M= moderate; H = high; N/A = not applicable because the facility had been closed at the time of data collection.

III. Evaluation of the Impacts of the Targeted Intervention: Year 5

A. Study Design

1. Sample selection process

Students targeted by this intervention are youth who are assigned to the care of the ODYS. These youth are eligible for Read 180 instruction at ODYS if: 1) assigned to the care of ODYS for more than six months; 2) determined to be "below proficient", but above "below basic" in Reading level as assessed by the Scholastic Reading Inventory (SRI); and, 3) if the youth is a non-high school graduate. Eligible youth are then split randomly between the treatment and comparison groups. Since there were youth in ODYS prior to the implementation of the project, the process of defining eligible youth and their random assignment will be discussed first followed by a description of this same process for those who were assigned to ODYS after the project began.

In August-September of 2006, all students in the care of ODYS were assessed using the SRI to determine their baseline Reading performance. The SRI assigned a Lexile score as a way of categorizing reading skill level, and any student that reads below grade level, but above "below basic" based on the SRI was eligible for assignment to the treatment condition. In ODYS, female students were allocated to one facility, and male students were allocated to one of the six male-only facilities based, in part, on the type of offense, available space, and programming needs. Eligible students were randomly assigned within each facility to the intervention or to the comparison condition using a computer-based random number generator specified by the evaluator. In addition, any student at grade level for reading was placed into the regular/traditional classroom. Thus, there are three groups of students: students in the intervention group in Read 180-only classrooms; students in the randomly selected comparison group that read below grade level based on Lexile scores; and students not assigned to either group because they read at or above grade level or "below basic" based on Lexile scores, or who have earned a high school diploma or a GED. The latter groups were together in the regular/traditional English classroom. Students who have graduated from high school or who have achieved their GED were not eligible for assignment. In order to populate the Read 180 classes, the initial random assignment to the Read 180 and traditional classes was made on a 60% - 40% allocation respectively. It should be noted that there are additional youth placed at ODYS who were beyond high school age, but below the age of 21 who were not enrolled in the high school program and therefore not part of the group under study.

For those youth assigned to ODYS after the initial allocation the selection process is as follows. Youth go through "intake", where they are processed and assessed for reading (using the SRI and the CAT) and for math (using the CAT) levels. Any youth that is eligible for the intervention based on the SRI is randomly assigned to either a Read 180 or to a traditional English class, but will attend traditional English classes at the "intake" facility until moved to his or her "home" facility. It is not until the youth is placed in their "home" facility that they will receive the Read 180 intervention, and then only if assigned to that intervention. The time between assignment to the Read 180 or traditional classroom and when the youth actually receives the intervention has been shown to be anywhere from 40 to 60 days and occasionally longer. Eligible students assigned to Read 180 or the traditional classroom after the initial 60-40 allocation, were assigned on a 50-50 allocation.

As students exit the ODYS, a "hole" is created in either the experimental/intervention or comparison/control condition. As new students are sentenced to the care of ODYS, they are assessed for eligibility, and randomly assigned to either the experimental or control group, if

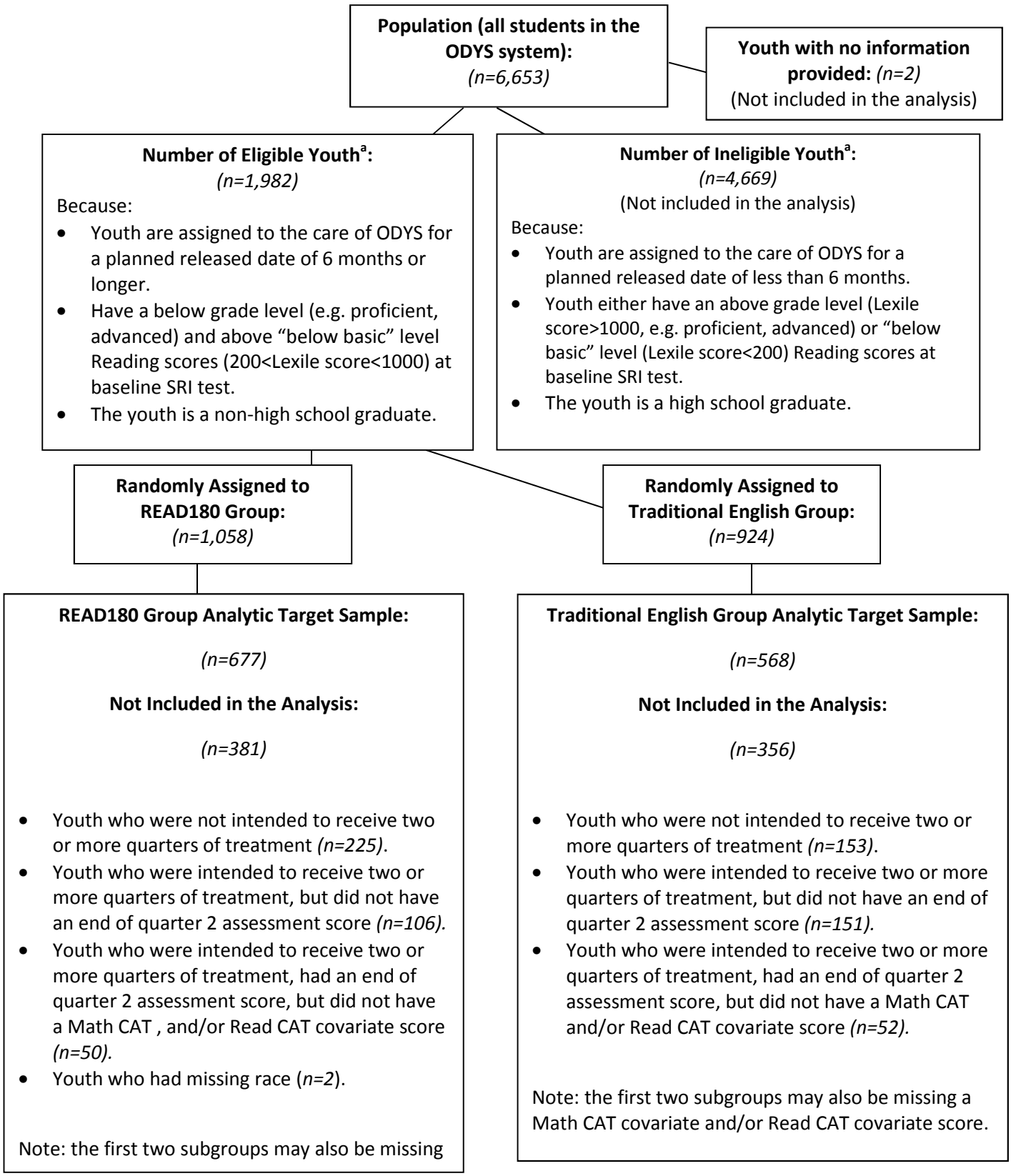
eligible. There is a limit of 15 students that can be assigned to any Read 180 class. This assessment and assignment procedure may have created minor glitches in the assignment of males to certain facilities and/or classes, but did not pose any problems for female assignment. Too, this method may have caused the number of students in a class at a given point in time to be less than 15, due to the fluid movement of youth between ODYS facilities or due to youth being removed from class due to disruptive behaviors. Exceeding the maximum number of fifteen youth in a Read 180 class was not an issue during the project.

2. Sample size

ODYS across five years has housed 6,653 youth. 1982 youth were identified as eligible for the targeted intervention with 1,058 (53%) assigned to Read 180 and 924 (47%) assigned to the traditional English classroom. This is the district wide sample size in the current report. The sample size for the HLM with the SRI as an outcome measure is 1,245 (677 Read 180 assigned, 568 traditional English assigned).

Since Scholastic makes the argument that only youth with two or more quarters exposure to Read 180 should be included in any impact analyses, youth who were not supposed to have any Read 180 treatment (they were in school for less than five weeks at any time during the first five years of the project) or who were supposed to have only one quarter of treatment, were omitted from the analyses. If data were missing (i.e., Lexile or CAT scores) for a given estimated model, the sample sizes presented here decreased further. See Figure 3 for all possible reasons for a decrease in sample size from the original random assignment sample size to the sample size used in the impact analyses.

Figure 1. Construction of the Years 1 through 5 Impact Sample from the Population: SRI as Outcome



[a] These youth have baseline SRI scores used as an indicator of their eligibility status

3. Power Analysis

In order to determine the probability of detecting real treatment effects, statistical power analyses were conducted for each ITT analytic sample in impact studies. Due to the nested data structure in this study, power analyses were guided by multi-level modeling research and IES guidelines (Hedges & Rhoads, 2010). Power was estimated based on the following realistic assumptions:

Two-level HLM model (student within school)
Type I error rate = 0.05, two-sided test
Intra-class correlation = 0.01
Treatment effect heterogeneity = 0.01
Amount of within-cluster variance explained by covariates = 0.30
Number of cluster-level covariates = 0

For each analytic group, the minimum effect size that could be detected at an acceptable power level of 80% and the power to detect an effect size of at least .33, were estimated respectively.

ANALYTIC SAMPLE 1: (Using SRI Lexile score after two quarters of intended treatment as the outcome, N = 1,245)
Number of clusters (schools) = 8
Number of individuals within each cluster = 155
Results:
Effect size = .132 Power = 80%
Effect size = .333 Power = 100.0%

ANALYTIC SAMPLE 2: (Using ReadCAT_1Year score as the outcome, N = 243)
Number of clusters (schools) = 7
Number of individuals within each cluster = 34
Results:
Effect size = .305 Power = 80%
Effect size = .333 Power = 85.7%

ANALYTIC SAMPLE 3: (Using ReadCAT_Last score as the outcome, N = 934)
Number of clusters (schools) = 7
Number of individuals within each cluster = 133
Results:
Effect size = .155 Power = 80%
Effect size = .333 Power = 100.0%

ANALYTIC SAMPLE 4: (Using SRI Lexile3_1YearCAT score as the outcomes, N = 225)
Number of clusters (schools) = 7
Number of individuals within each cluster = 32
Results:
Effect size = .315 Power = 80%
Effect size = .333 Power = 83.8%

ANALYTIC SAMPLE 5: (Using SRI Lexile3_LastCAT score as the outcome, N = 867)
Number of clusters (schools) = 7

Number of individuals within each cluster = 123
Results:
Effect size = .161 Power = 80%
Effect size = .333 Power = 100.0%

The above power analysis results indicate that all the analytic groups provide sufficient statistical power to detect a real treatment effect of one third standard deviation or even smaller sizes.

4. Description of the counterfactual

The randomly assigned comparison group received instruction in the traditional English classroom or resource room from a certified teacher. The traditional class period was between 45 and 55 minutes for a given day and had less time a week allocated to the class compared to 90 minutes of daily instruction for Read 180 students.

The student population in the traditional English classroom included those in the comparison group (e.g., eligible to participate in Read 180 program but assigned to the comparison group) and those not eligible, but still enrolled in school (e.g., read "below basic" and/or have a sentence of less than six months or have achieved "proficiency" on the SRI measure). Due to this unique population, many classes had students that were in different grades and operating at different academic and/or achievement levels.

Traditional classes are made up of youth at multiple grade levels, multiple disability levels, and multiple reading levels. For this reason, there is minimal group instruction and maximal individual and independent work being done. In the first two years of the project, youth came into the class at varying times, and had a folder geared to their learning level. When the majority of youth arrived to class, group instruction might have taken place, or there might have been an assignment on the board. Most teachers used assignments from the ODYS Central Office-issued text books for their subject area, and had multiple levels of these textbooks to accommodate the variety of learning levels that they would have encountered. While computers might have been used, it was normally for completion of projects, not for instruction.

In the last three years of the project American Education Cooperation's (AEC) A+ software was installed in all core subjects (e.g., history, mathematics, science and language arts). A+ is an interactive, research-based, curriculum software which customizes lessons based on each student's learning level. In language arts specifically students arrived to class, sat at a computer, logged in and began the day's lesson based on the prior day's progress. Students completed a variety of lessons, were assessed, and earned apples for their progress. The number of apples earned was the primary component of the student's grade in the class.

5. Data collection plan

There is a good, but arm's length relationship between the ODYS and the evaluation team at The Ohio State University (OSU). The staff at ODYS has been instrumental in helping the evaluation team gain timely entry into each of the youth facilities. They have also provided coded but de-identified data of each youth in the schools in a timely fashion on a quarterly basis. This occurred through ODYS personnel working at the State of Ohio Computer Center (SOCC). The ODYS staff, at the SOCC, supplied the evaluators with an electronic, encrypted, de-identified longitudinal data file containing student achievement, treatment assignment, daily class attendance, and student

movement records. Additional coded data were also provided on an as need or as available basis, e.g., listing of de-identified youth included in the Governor's early release program. Measures are categorized by (a) Youth measures and (b) teacher and classroom measures. These measures are described in more detail next.

Youth Measures. Data measuring student progress were collected by three means: 1) in the delivery of the specific intervention, 2) in the ODYS and ODE educational data systems, or (3) by the OSU evaluation team. Descriptions of these student measures are presented below.

(1) The SRI (Lexile score), a computerized, adaptive test that is used to assess reading level, is given as a pretest when youth first arrive at ODYS (e.g., at "in-take"). Youth are then reassessed quarterly while in the facility. If a youth is scheduled for release they will be assessed prior to their expected release date, if it is more than five weeks beyond the previous SRI assessment. This measure is utilized for eligible youth (in traditional or R180 classes) and ineligible youth (in traditional classes or recent graduates).

(2) The CAT in both reading and math is administered to all youth at intake. These tests, used to evaluate the youth's reading (vocabulary and comprehension) and mathematical achievement, are also given annually (at the end of the academic year; in spring quarter)[3] , provided that it is more than six months beyond the previous CAT assessment.

(3) The Ohio Graduation Test (OGT) is a state-wide achievement test administered to all youth in the State of Ohio initially in the 10th grade. This test has five components that cover reading, math, science, writing and citizenship. Students in the 10th grade at ODYS sit for the OGT. If a student is beyond the 10th grade and has not passed one or more sections of the OGT, they continue to sit for those sections of the test in the fall and spring of each year until they either pass that section(s) or leave the school system. No OGT analyses or results are presented in this report.

(4) Additional youth demographic characteristics are collected by the SOCC and given to the OSU evaluation team. They include: race, gender, disability status, degree obtained, degree expected, age, grade placement, chronological age grade placement, and special education status. In addition, the data provided by the SOCC also included daily attendance rosters for each youth in each class to be used to identify treatment amount as well as treatment of the treated and intent to treat groups.

(5) Students' Sense of Efficacy data have been collected by the OSU evaluation team. In the first two years of the project various efficacy measures were constructed and tested. In Years 3 and 4 a final Reading self-efficacy measure was administered. This survey was administered to each student entering DYS (at intake) and again at the end of the third and fourth year of the project. No student efficacy results are presented in this report.

Measures of teachers and classrooms. There are three central measures of classrooms: site visit classroom observations, classroom teacher/aide Read 180 implementation log, and teachers' sense of efficacy survey administration.

Classroom observations. An evaluation team member visits each school once per week during the instructional term. In the first two years of the project, the evaluator observed in one Read 180 classroom and at least one traditional classroom each week. In the third year of the project,

[3] In July 2007, staff at ODYS agreed that the CAT would continue as a student assessment tool until the end of the project.

classroom observations focused entirely on the Read 180 classroom observations. Depending on the year, the evaluation visits were designed to accomplish some of the following objectives:

1. Observe for the integrity and quality of instructional implementation of the Read 180 program

2. Observe for the components of the SIRI, Writing activities and HYS in the Traditional classes

3. Record the start time and rotation times of each observed class

4. Observe the climate of the building and classrooms

5. Observe for anomalies and idiosyncratic behaviors of teachers and students

6. Observe student participation, on-task behavior and student learning

7. Interact with classroom teacher, aide and literacy coach

8. Collect the weekly classroom implementation log for each Read 180 class

9. Administer and collect the student efficacy measures

10. Observe any skills taught in teacher professional development sessions

The Read 180 observation protocol form was initially supplied as a Scholastic Tool, however, minor modifications were made to fit the ODYS setting. In addition, specific tasks are looked for within each of the specified Read 180 rotations. For example, for whole groups, observers document whether students have their rBooks as well as whether the books are being utilized during classroom time. The Traditional Classroom observation protocol was not determined a priori and therefore was less specific and less detailed, but was made to relate with the Read 180 protocol whenever possible and became more structured as time progressed. Some examples of common fields include class start times, number of students, equipment used, length of group instruction, disruptive behaviors/removals, and number of aides present. For both Read 180 and traditional observations, each observer documents how much time is allocated to one-on-one instruction in the small group rotation.

In both winter 2009 and spring 2009, the two classroom observers observed three classrooms together as a means to assess inter-observer reliability. Across both quarters, the two observers were consistent on less than 30% of the Read 180 form. Therefore, in summer2009, a third evaluator went into the field to re-calibrate and re-train on key observation indicators. Given the level of inconsistencies, the quantitative observation data collected in Year 3 are not presented. However, the qualitative data gathered from the third evaluator in summer 2009 are presented when reporting the Year 3 observational findings. Two new classroom observers were hired in Year 4. After a quarter of training, these two evaluators observed twice a week. Inter-rater reliability across six quarters and across two years averaged 92%. Quantitative observation analyses are presented in the Year 5 implementation summary.

Further, data collection and cleaning issues across the two observers in Year 3 resulted in the dismantlement of the traditional English class collection of quantitative data at the end of Year 3. No observation data for the counterfactual will be presented in the Year 5 report.

Classroom Teacher/Aide Read180 implementation log. A log was created for the Read 180 teachers to maintain during the course of each 10-week Read 180 block. The purpose of this log is to capture the nature of the instruction as well as the degree of consistency and match between the paper curriculum and the actual reading curriculum. Data for each block included the actual amount of instruction occurring, an explanation of why the class was less than 90 minutes, if applicable, the number of minutes in whole group and wrap up as well as the minutes allocated to small group, individual learning and computer time for the first rotation.

Some content in the Read 180 observation protocol and implementation log did intersect. Specifically, the OSU observers recorded in their weekly observations the class start time (for the first three quarters), amount of instruction, and minutes allocated to rotations (for the last five quarters). These data were cross-validated with the data presented in the implementation log supplied by the classroom teacher/aide to determine the consistency between the information on the teacher log and the on-going Read 180 classroom practice.

Teachers' Sense of Efficacy Survey. A teachers' sense of efficacy instrument was pilot tested in the first two years of project implementation and administered at the end of spring 2009 in the third year, using all teachers across the seven ODYS facilities. Generally, teacher efficacy refers to teachers' confidence in their ability to bring about student learning and positive change (Ashton & Webb, 1986[4]). Since strong links have been found to exist between student achievement and teacher self-efficacy an assessment of teacher efficacy perceptions was thought to be useful (Gibson & Dembo, 1984[5]). The teachers' sense of efficacy instrument consisted of three pre-existing teacher efficacy instruments – the Teachers' Sense of Efficacy Scale (TSES) by Tschannen-Moran & Woolfolk Hoy, 2001[6], the Teacher Efficacy Scale (TES) by Gibson & Dembo, 1984, and Collective Efficacy Scale (CES) by Goddard, Hoy & Woolfolk Hoy, 2000[7]. In Year 4, the teachers' sense of efficacy instrument was modified. The Gibson and Dembo items were removed and additional items measuring school climate were added. Given that the scores were found to lack construct validity, no teacher perceptions on efficacy and climate were presented in the Year 4 report. No teacher efficacy data were collected in Year 5.

6. Summary of analytic approach to the impact analysis

Models. Four primary impact models were estimated: (1) a cross sectional Intent-To-Treat (ITT) hierarchical linear model with SRI after two quarters of treatment as the primary outcome, (2) a cross sectional ITT hierarchical linear model with Read CAT after one year at ODYS as post assessment as the primary outcome, (3) a cross sectional ITT hierarchical linear model with the youths last Read CAT as the primary outcome, and (4) a longitudinal ITT hierarchical linear model with SRI as the primary outcome. Appendix A presents the estimated models in more detail (see Appendices A3 and A4 for the descriptive statistics and estimated models with more detailed results). Tests of equivalency were conducted (results presented in appendix 7) to ensure that

[4] Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. New York: Longman.

[5] Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 569-582.

[6] Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching & Teacher Education: An International Journal of Research and Studies, 17*, 783-805

[7] Goddard, R. D., Hoy, W. K., & Woolfolk Hoy, A. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*(2), 479-507

youth who were in the HLM analyses were not statistically different from those who were omitted and to ensure comparability of treatment groups at baseline. The youths were not statistically different on the baseline outcome measures, demographic characteristics or included or excluded from the outcome analyses.

Intent to treat (ITT) was defined in this study differently than it has been in more conventional experimental studies. Intent to treat traditionally has been defined by the length of the project, however, youth mobility in and out of the facility makes it a challenge to define ITT based on program start and end date. Therefore, intent to treat was defined by each youth's entrance into ODYS and exit out of the facility. Appendix A1 describes in more detail the methods by which ITT youth were defined.

Model specifications. Hierarchical linear modeling (HLM) was used to evaluate the overall targeted intervention impact of the Striving Readers Initiative on the Reading performance of the low-achieving incarcerated youth due to its methodological advantages (Raudenbush & Bryk, 2002[8]; Singer & Willett, 2003[9]). The impact studies focused on the Intent-To-Treat (ITT) youth (i.e., those who had the opportunity to receive the treatment). Two-level HLMs were used for the cross-sectional analyses to account for the clustering effect and multiple student characteristics. Multiple linear regressions were also fitted under the circumstances where there was no between-school variance. In addition, a longitudinal analysis of the repeated measures of SRI was also conducted using HLM for the ITT sample, and the relevant results are presented at the end of this section.

Since Scholastic makes the argument that only youth with at least two quarters' exposure to READ 180 should be included in any impact analyses, youth who were not supposed to have any READ 180 treatment (they were in school for less than five weeks at any time during the first four years of the project) or who were supposed to have only one quarter of treatment, were omitted from the ITT analyses.

Note that for all ITT analyses, list-wise deletion was used to remove subjects with missing data from each analytic sample. All covariates, with exception of the treatment predictor, were grand mean centered. Covariates with p values of .200 or above in the full models were not included in the parsimonious final models. Appendix A4 presents the specification of the models and the detailed results.

B. Description of the First-, Second-, Third-, Fourth-, and Fifth- Year Sample

1. Basic characteristics of teachers

The ODYS intervention staff has a healthy representation of teaching experience. All of the Scholastic Read 180 teachers are English/Language Arts certified and all of the teacher aides have proper certification. Table 7 shows characteristics of the teaching staff, which include their start date, end date, (if applicable), gender, teaching experience and degree attainment. Of the seven teachers and seven aides, one of the teachers and four of the aides were existing ODYS employees.

[8] Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage publications, Inc.

[9] Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis*: Modeling change and event occurrence. NY: Oxford.

Starting in Year 3, vacancies became more prominent. Three facilities had a literacy coach position vacant for three months (Facility 7), six months (Facility 1) and almost a year (Facility 5). Facility 4 had a teacher position vacant for three months. Teachers hired in year 3 tended to have slightly less teaching experience than those teachers that there were replacing. A new Read 180 teacher was hired in Facility 8 after the veteran teacher retired.

2. Basic characteristics of classrooms

The Read 180 classroom is carpeted with 5 computer stations and headphones, a reading area with couches and books to select based on personal preference and reading level, and tables arranged in a group or groups, depending on the size of the classroom. It is a highly-structured class, with the first 20 minutes of whole group being conducted with all of the class, then splitting into smaller groups for 20 minutes each of computer work, independent reading, and small group. The model calls for a 10 minute wrap up with the whole group at class-end, but this did not occur for the majority of the first three terms in year 3 due to the movement issues previously described. The whole group rotation continued to be omitted in Years 4 and 5. Each Read 180 classroom has a teacher and an aide, and access to the Literacy Coach.

In contrast with this are the typical traditional English (and most other) classes, where youth do individual work that in the first two years of the project was many times previously assigned and kept in folders, with the teacher giving help as needed. In the third year of the project, and after the implementation of A+ individual work is still central but instead of the use of the worksheets and text books, students focus attention on computer-generated lessons.

Sometimes group work is done, but most times this is not practical because a typical class will have reading levels ranging from fourth to twelfth grade in addition to having students with disabilities. Most traditional English classrooms have 8-15 students without an aide or additional help. Classes are typically unstructured with little or no group instruction, and no room or materials for independent reading. There is, however, a library that students have access to, and some of the teachers bring in outside materials that are relevant to the subject being taught, so that the youth may have access to other material.

Table 7. Current and Past Teacher Characteristics by Facility

	Facility 1	Facility 2	Facility 3 - Closed	Facility 4	Facility 5	Facility 7	Facility 8
Name	AJ	KH	KM	KH	AV	SD	LS
Start Date	7/30/2007	9/5/2006	11/13/2006	3/2/09	8/7/2006	12/1/2008	1/6/2010
End Date	Current Teacher*	Current Teacher	Current Teacher*	Current Teacher	Current teacher	Current Teacher	Current Teacher
Experience	3 yrs - sub; 1 yr - tutoring contract; 2 yrs teaching under contract	12 years	2 year	12 years	6 yrs subbing, 1 yr Cols Public Schools, 2.25 yrs DYS	1 year	6 years
Gender	Female	Female	Female	Female	Female	Male	Female
Degree obtained	Master of Arts in Teaching Secondary Education	Bachelor of Science in Education (Currently working on Master in English Composition)	BS Univ. of Akron: from the license: (63) Adolescent to Young Adult (ages 12-21/grade 7-12: 050145 Integrated Language Arts)	BS - University of Akron Reading k-12 Elementary 1-8	OSU - English Degree, Ohio Dominican - Teacher Licensure in Integrated Language Arts 7-12	MA – English Education – Morehead State University	BA – Social Welfare MA – English and Lit. Education
Name-Teacher	SM		SK	JB		AV	CM
Start –End Date	06/25/06-7/27/2007		9/4/2006- 11/13/2006	10/1/08-2/27/09		6/25/2006-12/1/2008	7/22/2007-12/20/2009
Experience	2 yrs		25 yrs	30+ years		6 years	30 years
Gender	Female		Female	Female		Female	Female
Degree obtained	BS - English Ed		BS in Comprehensive Communications	BS - English		Bachelor	BS - English
Name-Teacher				KK			ACG
Start Date				9/3/2006			8/21/2006-6/8/2007
End Date				10/1/2008			5 yrs
Experience				8 years			Female
Gender				Female			BS - English Ed
Degree obtained				College of Wooster-BA-May,1999; Univ of Arizona - M. Ed. May, 2008			ACG

*current at the time of facility closure

3. Basic characteristics of students

Demographic information for students serviced at ODYS in the five years of the targeted intervention component of the project is presented in Table 8. For both Read 180 and traditional English groups, the primary racial category is Black (70.3% for Read 180 and 68.2% for traditional English group), followed by White (22.9% and 25.7% respectively). The majority of the students who are eligible for treatment (96.2%) are male and only a small portion of them (3.8%) are female.

Half of the incarcerated youth have disability status (50.2% and 46.3% respectively) and are classified as special education (44.8% and 42.4% respectively). When a disability exists, it is primarily Emotional Disturbance (20.5% and 19.2% respectively), followed by Specific Learning Disability (16.9% and 15.2% respectively). There is some representation of Cognitive Disabilities in the youth (8.6% for both groups). Most of the youth are 18-22 years old (as of Dec 2011; students could have been up to 5 years younger than the calculated age if they had been enrolled in the program at the begininning in 2006), with a portion of them under age 18 (9.7% Read 180 and 11.3% traditional) and above age 22 (9.1% and 6.2% respectively). Around 30% of them have attained a ninth grade academic status, and around 25% have a tenth grade status. In addition, approximately 25% of the Read 180 and traditional English youth have graduated. Graduation percentages here are slightly misleading as youth could have been housed in ODYS exposed to either Read180 or Traditional English instruction in dosage variations and then exited ODYS. The youth could have earned their diploma in their home town, or graduation status of the youth was unknown but the youth was beyond graduation age and therefore forced into the graduation category.

Table 8. Demographic Descriptions Disaggregated by Treatment Group Across Five Years

Demographic Category		R180		Traditional	
	Demographic Option	Freq	%	Freq	%
Race	Asian	1	.1	0	0
	Black	586	70.3	526	68.2
	Hispanic	19	2.3	14	1.8
	Native American/Alaskan	1	.1	2	.3
	White	191	22.9	198	25.7
	Multiracial	34	4.1	30	3.9
	Missing	1	.1	1	.1
Gender	Male	801	96.2	742	96.2
	Female	32	3.8	29	3.8
Special Education	No	460	55.2	444	57.6
	Yes	373	44.8	327	42.4

Table 8. Demographic Descriptions Disaggregated by Treatment Group Across Four Years (continued)

		R180		Traditional	
Demographic Category	Demographic Option	Freq	%	Freq	%
Disability Status*	Au	2	.2	1	.1
	CD(MR)	72	8.6	66	8.6
	Df	1	.1	0	0
	ED	171	20.5	148	19.2
	MD	5	.6	4	.5
	O-Min	18	2.2	18	2.3
	O-Maj	1	.1	0	0
	OI	1	.1	0	0
	SL	4	.5	1	.1
	SLD	141	16.9	117	15.2
	TBI	1	.1	1	.1
	Vi	1	.1	1	.1
	Non-disabled	415	49.8	414	53.7
Age **	15	2	.2	5	.6
	16	23	2.8	18	2.3
	17	56	6.7	65	8.4
	18	98	11.8	104	13.5
	19	169	20.3	143	18.5
	20	155	18.6	173	22.4
	21	137	16.4	123	16.0
	22	117	14.0	92	11.9
	23	58	7.0	41	5.3
	24	12	1.4	5	.6
	25	6	.7	2	.3
Current Grade	8	4	.5	6	.8
	9	214	25.7	219	28.4
	10	230	27.6	186	24.1
	11	118	14.2	97	12.6
	12	57	6.8	61	7.9
	13***	210	25.2	202	26.2

Note: The disability status acronyms include: Au = Autism; CD(MR) = Cognitive Disability-Mental Retardation; Df = Deafness; Ed = Emotional Disturbance; MD = Mental Retardation; O-Min = Other Impairment-Minor; O-Maj = Other Impairment-Major; OI = Orthopedic Impairment; SL = Speech or Learning Disability; SLD = Specific Learning Disability; TBI = Traumatic Brain Injury; VI = Visual Impairment.

* If a person was categorized as being disabled, this is his/her disability type.

**Age was calculated by taking 2011 and subtracting the year in which the youth was born. Youth could be as much as 5 years younger than the calculated age at the time they received treatment.

*** If youth did not have a graduation status but had left DYS or if they were in the appropriate age to graduate, they were forced into grade 13. Some in the grade 13 have actually graduated.

C. Impacts on Students at the End of Five Years

Hierarchical linear modeling (HLM) was used to evaluate the overall targeted intervention impact of the Ohio Striving Readers Initiative on the reading performance of the low-achieving incarcerated youth due to its methodological advantages (Raudenbush & Bryk, 2002[10]; Singer & Willett, 2003[11]). The impact studies focused on the Intent-To-Treat (ITT) youth (i.e., those who had the opportunity to receive the treatment). Two-level HLMs were used for the cross-sectional analyses to account for the clustering effect and multiple student characteristics. Multiple linear regressions were also fitted under the circumstances where there was no between-school variance. In addition, a longitudinal analysis of the repeated measures of SRI was also conducted using HLM for the ITT sample, and the relevant results are presented at the end of this section.

Since Scholastic makes the argument that only youth with at least two quarters exposure to Read 180 should be included in any impact analyses, youth who were not supposed to have any Read 180 treatment (they were in school for less than five weeks at any time during the entire five years of the project) or who were supposed to have only one quarter of treatment, were omitted from the ITT analyses.

Note that for all ITT analyses, list-wise deletion was used to remove subjects with missing data from each analytic sample. All covariates, with exception of the treatment predictor, were grand mean centered. Covariates with p values of .200 or above in the full models were not included in the parsimonious final models. Appendix A presents the specification of the models; the results with Appendix A7 addressing test of equivalency for the cross-sectional HLM with the SRI as outcome.

Table 9: Estimated Impact of Targeted Intervention on SRI Lexile Outcome of ITT Incarcerated Youth after Two Quarters of Intended Treatment Aggregated across Five Years of the Project Data

Population Group	Unadjusted Means		Regression-Adjusted Means		Estimated Impact	Effect Size	*p* Value	*Power (MDES)*
	Control	Treatment	Control	Treatment				
All ITT incarcerated youth across five years	798.52	840.38	791.69	850.83	59.14	0.21	<.001	0.12

In ODYS, SRI has been serving as the major test instrument for incarcerated youth. For each youth, the SRI was taken at baseline, and then repeated at the end of each academic term. Thus the Lexile scores of the ITT youth after being offered two quarters of treatment were used as the outcome measure in the first cross-sectional impact study. A final ITT sample of 1,245 youth across the entire five years of the project was included in this analysis.

As seen in Table 9, the analysis detected that the Read 180 program had a significant overall impact on the low-achieving youth's SRI Lexile outcome. Youth in the Read 180 group on average performed 59.14 points higher than their comparison counterparts after being offered two quarters

[10] Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage publications, Inc.

[11] Singer, J. D. & Willett, J. B. (2003). *Applied longitudinal data analysis*: Modeling change and event occurrence. NY: Oxford.

of treatment. The effect size measured by Glass's delta (0.21) was fairly substantial given the huge variability of Lexile scores.

Table 10. Estimated Impact of Targeted Intervention on ReadCAT Outcome of ITT Incarcerated Youth Aggregated across Five Years of the Project Data

Population Group	Unadjusted Means		Regression-Adjusted Means		Estimated Impact	Effect Size	*p* Value	*Power (MDES)*
	Control	Treatment	Control	Treatment				
ITT youth with ReadCAT_Last[a] score across five years[*]	6.45	6.69	6.44	6.69	0.25	0.09	0.106	0.16
ITT youth with ReadCAT_1Year[b] score across five years	5.63	6.06	5.58	6.19	0.61	0.26	0.011	0.28

[*] The results based on multiple linear regression were presented for the analysis sample.
[a] The last available record for post-test of ReadCAT
[b] ReadCAT score after one year of intended treatment

Since the SRI test is more often practiced by the youth in READ 180 and its psychometric properties are not well-established for the targeted population, using a second outcome measure in the impact study is especially important in the evaluation of this initiative. The ReadCAT is an additional reading assessment that has been administered to the ODYS youth. The ReadCAT variable is a grade level equivalent metric. In addition to the baseline measure, post-test scores of ReadCAT were also available for some ITT youth. Unlike the SRI measure that is administered at the end of each quarter, the ReadCAT is generally administered at the end of each academic year (usually at the end of spring term). Unfortunately, even for youth who were frequently tested by ReadCAT after baseline, they did not have consistently timed test data mainly because either the institution did not assess students regularly using Read CAT or the youth was not housed at DYS at time of assessment. The administration of subsequent ReadCAT assessments generally occurred at the end of spring term, however the level of adherence to this schedule varied across institutions. Thus the elapsed time between different ReadCAT administrations differed greatly (e.g., a month, half a year, more than a year, etc.). This issue presented a major challenge in obtaining a cleaned measure of reading using ReadCAT as an outcome measure.

Two approaches were employed to generate the ReadCAT outcome for the cross-sectional ITT analyses. We first obtained the last available record of post ReadCAT scores from each subject and used it as the outcome variable in the impact study. Thus a subject must have at least one post measure of ReadCAT to be included in the analysis. A total of 934 ITT youth across the five years of the project were included in this study. According to Table 10, this analysis did not find any significant overall impact of the READ 180 program on the low-achieving incarcerated youth based on their last post-test score of ReadCAT. In this analysis, the READ 180 youth only had a slightly higher mean scale score than the youth in the comparison group(6.69 vs. 6.44), and the effect size was small (0.09).

While the first approach provided us with the largest possible sample size for the outcome analysis of ReadCAT, a major concern about this analysis was the mistimed test data as mentioned

previously. Therefore, a second approach was used to obtain a cleaner outcome measure of ReadCAT: based on the test administration dates, the post-test ReadCAT score measured within an approximate time interval of one year (the exact time interval used in data cleaning was 365 days ± 60days = [305 days, 425 days]) from baseline was selected as the outcome in a separate impact analysis; if more than one post-test score was available for a given subject, the one nearest to 365 days was retained. This rule yielded a much smaller final study sample of 243 youth across the five years. The results in Table 10 indicated that different from the first ReadCAT analysis, there was a significant overall impact of the READ 180 program on the low-performing incarcerated youth based on their ReadCAT scores measured after approximately one year of supposed treatment. The treatment youth outperformed the comparison youth by an average of 0.61 scale points, with an effect size of 0.26.

Appendix A6 presents more information for the two ReadCAT analysis samples. It can be seen that the last available post record of ReadCAT is a much messier measure than the other ReadCAT outcome generated by the second approach. The amount of time between baseline and last ReadCAT score could range from no more than 1 quarter to approximately 16 quarters. The average length of stay in DYS is approximately 10.5 months, and typically the longer the target youth are in control of DYS, the more severe their felonies are. For youth who had a very long length of elapsed time between their baseline and the last available post ReadCAT, there is a great possibility that these youth were first incarcerated, then released, and were readmitted to DYS because of recidivism. Note that in some cases, the time lapse was as long as three to four years and the mean gain scores for those youth would have little to do with the program impact of Read 180. All these confounding factors may explain why we found no significant result when using the last available post measure of ReadCAT as the outcome. Note that there was also a substantial difference between the sample sizes of the two ReadCAT analyses and the percent of overlapping subjects belonging to the same time lapse intervals in each respective sample is quite small. Therefore, one may want to rely more on the results based on the much cleaner outcome, the post ReadCAT measured after approximately one year of intended treatment.

Also note that for the first analysis using the last available post measure of ReadCAT as the outcome, the HLM was initially fitted to the data but it turned out that the between-school variance was zero. Thus multiple linear regression analyses were refitted to the data and resulted in the same regression coefficients obtained by HLM. The second ReadCAT analysis did not encounter this problem so the HLM coefficients were reported in Table 10.

In the previous impact analyses, conclusions were different depending on whether the outcome measure was SRI or different post-test measures of ReadCAT. To further confirm the consistency of the findings, additional analyses of the SRI Lexile scores were conducted for the two analytical samples using ReadCAT as the outcome: for each analysis sample, the corresponding Lexile score obtained after two quarters of supposed treatment by each subject was used as the outcome measure. Due to missing data, some subjects who were included in the ReadCAT analyses were dropped from these two parallel analyses. A total of 867 ITT youth were included in the first parallel impact analysis, and 225 in the second one.

Table 11. Estimated Impact of Targeted Intervention on SRI Lexile Outcome of ITT Incarcerated Youth based on ReadCAT_Last Analysis Samples Aggregated across Five Years of the Project Data

Population Group	Unadjusted Means		Regression-Adjusted Means		Estimated Impact	Effect Size	*p* Value	*Power (MDES)*
	Control	Treatment	Control	Treatment				
ITT youth with Lexile3_LastCAT[c] score across five years[*]	776.72	836.26	771.39	840.59	69.20	0.25	<.001	0.15
ITT youth with Lexile3_1YearCAT[d] score across five years[*]	760.40	809.12	745.29	820.16	74.87	0.28	0.006	0.28

[*] The results based on multiple linear regression were presented for the analysis sample.

[c] The SRI Lexile score measured after two quarters of supposed treatment for the analysis sample who had the last available post-test record for ReadCAT as the outcome

[d] The SRI Lexile score measured after two quarters of supposed treatment for the analysis sample who had a post-test score of ReadCAT after approximately one year of intended treatment as the outcome

Based on Table 11, both parallel analyses detected a significant overall impact of the READ 180 program on the Lexile outcome of the low-performing incarcerated youth. The significant findings were consistent with those found in the first cross-sectional impact analysis using the SRI Lexile outcome (see Table 9) and those found in the analysis based on the post-test ReadCAT scores measured after approximately one year of supposed treatment (see Table 10).[12] In addition, the magnitude of the effect sizes in these two parallel analyses were quite similar to the previous two analyses with significant findings, which were substantially larger than the analysis using the last available post record of ReadCAT as the outcome (see Table 10).

Note that for both parallel analyses using the SRI Lexile as the outcome, the HLM analysis was first attempted but encountered the same problem with the between-school variance as before in the analysis of last available post measure of ReadCAT, so multiple linear regression analyses were used again and generated the same regression coefficients obtained by HLM.

Additional Analysis. Since the project also involved a longitudinal design, an analysis of repeated measures of the SRI was also of interest in this evaluation. Therefore, a longitudinal HLM analysis was carried out for the 1,393 ITT youth who had at least one post measure of the SRI in addition to the baseline. A total of 7,334 observations across 21 possible time points[13] (i.e., baseline + 4 * 5) were included in the analysis.

[12] Note that the impact estimates for these two parallel Lexile analyses were approximately 70 points or higher, which was slightly larger than the estimated impact generated by the first cross-sectional SRI Lexile analysis (about 60 points) based on the overall ITT sample.

[13] According to the longitudinal data, the maximum number of SRI repeated measures obtained for a subject was 16.

Table 12. Estimated Fixed Effects in the Final Linear Longitudinal Model Based on SRI Lexile Scores Aggregated across Five Years of the Project Data

Fixed Effects		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2
Intercept	α_0	788.26	6.863	114.85	<.001	--
White	α_1	-17.43	11.557	-1.51	0.132	0.00
Age	α_2	5.70	2.776	2.05	0.040	0.00
Base_MathCAT	α_3	8.96	2.317	3.87	<.001	0.01
Base_ReadCAT	α_4	37.16	2.392	15.54	<.001	0.15
Disability	α_5	-44.78	9.561	-4.68	<.001	0.02
Grade Level	α_6	16.57	3.123	5.31	<.001	0.02
Mobility	α_7	-2.87	9.419	-0.31	0.760	0.00
TRTGroup	α_8	3.40	9.342	0.36	0.716	0.00
Time	β_0	0.10	2.477	0.04	0.969	0.00
White*Time	β_1	10.00	4.033	2.48	0.013	0.01
Age*Time	β_2	-4.62	0.954	-4.84	<.001	0.03
Base_ReadCAT*Time	β_3	1.90	0.700	2.71	0.007	0.01
Mobility*Time	β_4	9.70	3.354	2.89	0.004	0.01
TRTGroup*Time	β_5	19.56	3.266	5.99	<.001	0.04

As shown in Table 12, it was found that READ 180 had a significantly positive longitudinal impact on the SRI Lexile outcome of low-performing incarcerated youth, with a constant growth rate over time. Specifically, compared to the youth instructed by the traditional English class, the students in READ 180 on average gained 19.56 more Lexile points after each term, while controlling for other covariates, with an effect size (measured by Cohen's f^2) of 0.04.

In addition, the results indicated that the baseline scores of CAT (both Reading and Math) and a few demographic variables (e.g., age, disability, and grade level) were statistically significant in the final growth model, explaining some variability in the initial Reading status and/or the Reading growth rate of the low-achieving incarcerated youth.

Appendix A: Impact Analysis Methods

Appendix A1. Defining TTT and ITT Groups based on a minimum of 5 weeks of treatment received for each quarter

First, youth were identified based on the amount of treatment they were supposed to have received. To identify a youth with respect to ITT, the eligible youth was first categorized based on the amount of treatment received in each quarter, without regard to whether or not they should receive treatment in other quarters. Youth were categorized across the eight possible quarters as receiving: (a) two quarters of treatment, (b) three quarters of treatment, (c) four quarters of treatment, and so on (identified as treatment amount in future analyses). Youth were categorized into these groups if they attended at least half of the quarter's class session. Notably, youth could receive treatment in any possible quarter combination (i.e., two quarters of treatment in Fall and Summer quarters, or two quarters of treatment in Spring and Summer quarters).

Youth were then compared against how many classes they were supposed to have attended. Intent to receive treatment for the eligible, traditional English assigned youth is identified by assignment date. Read180, assigned, intent to treat youth are identified by their classroom placement date. If a youth was identified in the first five weeks of a given quarter as either being assigned to the traditional English class (comparison group) or actually in the treatment classroom (Read 180 group) they were classified as intent to treat in that given quarter. If a youth was assigned/placed in their designated classroom in the 6th week of the quarter or after they are classified as intent to treat for the next quarter.

If a youth never left ODYS and/or the school system the youths amount of treatment was compared to when he or she was eligible to receive at least five weeks of treatment. For example, Youth A was placed in Read 180 in September 1, 2006, this youth was eligible to received Read 180 treatment in the first quarter of the project. He never left the facility and therefore should have received eight quarters of treatment. If he received those eight quarters, that is, attended at least five weeks a quarter of Read 180 sessions for each of the eight quarters, he was identified as treatment of the treated, otherwise, he was identified as intent to treat but not treated.

It is possible that youth who are in good standing will be released early by the juvenile court judge, and this may substantially decrease the amount of Read 180 or English classroom treatment a youth receives. Further, a youth can earn his or her GED or high diploma and no longer be enrolled in high school classes (but still be housed at ODYS). If a youth left school, his or her intent to treat status stopped. For example eligible Youth B was randomly assigned to a traditional class on May 20th, 2007 and was subsequently identified as intent to treat in the 4th quarter of the project. He then was released from ODYS on March 10th, 2008. He was supposed to have three quarters of treatment. This is compared to how much treatment he actually had. If he had three quarters of treatment then he was identified as treatment of the treated. Otherwise, he was identified as ITT. This latter issue often happens when a youth refused to attend class or was penalized in lock down for disruptive behavior.

Finally youth can be released from ODYS only to return months or years later. Again, ITT identification is defined by when they were housed at ODYS. If a youth was placed in Read 180 for example and left ODYS, and then came back, only the youth's time in the facility was counted towards ITT. Take as an example Youth C. She was placed in Read 180 October 15th, 2006, left ODYS on February 15th, 2007, and arrived back at ODYS on November 5th, 2008. Her first stay she was supposed to receive two quarters of treatment and three more quarters of Read 180 in her second stay. She was identified as five quarters of intent to treat. If she received five quarters then she was identified as treatment of the treated. If she received less than five quarters then she is identified as intent to treat but not treated.

Appendix A2. Selection of covariates

A substantial amount of decision making and cleaning was needed to ensure this score was a viable covariate. There were two general issues associated with the collection of Read/Math CAT scores. First, there were roughly 500 youth who had a baseline test taken before they were recorded as entering ODYS. This either means an error in the data file or these youths were assigned to ODYS prior to their first baseline assessment but this entrance date was not recorded in the file provided to the OSU evaluation team. Second, the 500 youth just mentioned as well as other youth in the data file (roughly an additional 1,000 youth) took the their base line test prior to August 1, 2006, with many youth taking the test as early as 2000. Given the age sensitivity of this assessment we believed it was problematic to use their first test score as the baseline score without further investigating how we might circumvent this problem. Therefore, a series of decision rules were developed. The following rules were applied in cleaning the CAT scores. These rules are as such:

1) If the youth has a score that is prior to July of 2006, has been at the facility at time of project implementation, and has another score two months prior to or up to the date of program implementation, then the latter score was utilized as the baseline test.

2) If the youth has a score that is prior to July of 2006 but came to the facility after project implementation (e.g., Winter 07 or after), the test that was administered up to two months after their arrival was utilized as the covariate. This decision was made given the fact that, as previously discussed, after the first quarter there was an average 40-60 day turn around to place youth in the classroom. Therefore, we believe that waiting two months will not negatively effect the youth's baseline assessment since it is unlikely they would have received treatment during this time span.

3) If the youth only has one CAT score and it is out-of-date, then the date that the test was administered will determine if it is used as a covariate. That is, if a score was administered after three months of arriving to the facility or assessed July 2005 or before if at ODYS when the project began, such a score will be treated as missing.

4) Finally, a case by case decision for the appropriate CAT covariate was made for those youth who were released from ODYS and subsequently returned. Attention was given to when the test was administered (before or after July 2005) and to the test administration date that is closest to the second time they arrived at ODYS.

Overall these rules were implemented to ensure the covariate score utilized came as close to when the youth first was introduced to the Read 180 material (if assigned to Read 180) or close to the start of the project or entrance to ODYS (if assigned as ineligible or assigned to traditional English).

Appendix A3. Targeted Intervention Descriptive Statistics

Table A3. 1. Summary Statistics of SRI Lexile Outcome after Two Quarters of Treatment for Targeted Intervention ITT Analysis Sample Across Five Years of Data

Analysis Sample	Group	Mean	SD	School Sample Size	Student Sample Size
All ITT incarcerated youth across five years (Sections 1 & 2)	Control	798.52	280.96	8	568
	Treatment	840.38	264.48	8	677
	Total	821.28	272.81	8	1245

Table A3. 2. Summary Statistics of ReadCAT Outcome for Targeted Intervention ITT Analysis Samples Across Five Years of Data

Analysis Sample	Group	Mean	SD	School Sample Size	Student Sample Size
ITT incarcerated youth with ReadCAT_1Year score across five years (Sections 3 & 4)	Control	5.63	2.37	7	110
	Treatment	6.06	2.50	7	133
	Total	5.87	2.44	7	243
ITT incarcerated youth with ReadCAT_Last score across five years (Sections 7 & 8, including A & B)	Control	6.45	2.74	7	430
	Treatment	6.69	2.77	7	504
	Total	6.58	2.76	7	934

Table A3. 3. Summary Statistics of SRI Lexile Outcome Associated with ReadCAT ITT Analysis Samples Across Five Years of Data

Analysis Sample	Group	Mean	SD	School Sample Size	Student Sample Size
ITT incarcerated youth with Lexile3_1YearCAT score across five years (Sections 5 & 6, including A & B)	Control	760.40	270.91	7	95
	Treatment	809.12	243.14	7	130
	Total	788.55	255.79	7	225
ITT incarcerated youth with Lexile3_lastCAT score across five years (Sections 9 & 10, including A & B)	Control	776.72	273.85	7	389
	Treatment	836.26	258.90	7	478
	Total	809.54	267.20	7	867

Appendix A4. Targeted Intervention Estimated Models

Section 1: Full Hierarchical Linear Model for Cross-Sectional ITT Analysis of SRI Lexile3 Across Five Years of Data

For student *i* in institution *j*,
Level 1:

$$\begin{aligned}\text{LEXILE2}_{ij} = {} & \alpha_{0j} + \alpha_{1j}(\text{LEXILE0}_{ij} - \overline{\text{LEXILE0}}..) + \alpha_{2j}(\text{WHITE}_{ij} - \overline{\text{WHITE}}..) + \alpha_{3j}(\text{AGE}_{ij} - \overline{\text{AGE}}..) \\ & + \alpha_{4j}(\text{MATHCAT}_{ij} - \overline{\text{MATHCAT}}..) + \alpha_{5j}(\text{READCAT}_{ij} - \overline{\text{READCAT}}..) + \alpha_{6j}(\text{DISB}_{ij} - \overline{\text{DISB}}..) \\ & + \alpha_{7j}(\text{GRDLVL}_{ij} - \overline{\text{GRDLVL}}..) + \alpha_{8j}(\text{MOBL}_{ij} - \overline{\text{MOBL}}..) + \alpha_{9j}(\text{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$
$$\alpha_{7j} = \alpha_{70}$$
$$\alpha_{8j} = \alpha_{80}$$
$$\alpha_{9j} = \alpha_{90}$$

Table A4. 1. Fit Indices for the Full Cross-Sectional Model: Cross-Sectional ITT Analysis of SRI Lexile3 Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	16906.4	16930.4	16931.4

Table A4. 2. Estimated Fixed Effects in the Full Cross-Sectional Model: Cross-Sectional ITT Analysis of SRI Lexile3 Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	792.0300	13.5680	58.38	<.0001	--	--
Lexile0	α_{10}	0.5216	0.0396	13.18	<.0001	0.14	0.00
White	α_{20}	2.0453	15.5769	0.13	0.8956	0.00	0.01
Age	α_{30}	-11.1889	3.9009	-2.87	0.0042	0.01	-0.04
MathCAT	α_{40}	7.0491	3.3996	2.07	0.0383	0.00	0.03
ReadCAT	α_{50}	27.3037	3.4525	7.91	<.0001	0.05	0.10
Disability	α_{60}	-14.6606	13.7149	-1.07	0.2853	0.00	-0.05
Grade Level	α_{70}	12.6705	4.5250	2.80	0.0052	0.01	0.05
Mobility	α_{80}	14.8715	12.4645	1.19	0.2331	0.00	0.05
TRTGroup	α_{90}	59.6962	12.2684	4.87	<.0001	0.02	0.21

Table A4. 3. Estimated Random Effects in the Full Cross-Sectional Model: Cross-Sectional ITT Analysis of SRI Lexile3 Across Four Years of Data

Variance Component (Full Model)	Estimate	SE	*z*-value	*p*-value
σ^2	45954.00	1848.50	24.86	<.0001
τ_{00}	607.87	587.90	1.03	0.1506
Variance Component (Unconditional Model)				
σ^2	73593.00	2958.43	24.88	<.0001
τ_{00}	381.10	491.85	0.77	0.2192

Section 2: Final Hierarchical Linear Model for Cross-Sectional ITT Analysis of SRI Lexile3 Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\begin{aligned}\text{LEXILE2}_{ij} = {} & \alpha_{0j} + \alpha_{1j}(\text{LEXILE0}_{ij} - \overline{\text{LEXILE0}}..) + \alpha_{2j}(\text{AGE}_{ij} - \overline{\text{AGE}}..) \\ & + \alpha_{3j}(\text{MATHCAT}_{ij} - \overline{\text{MATHCAT}}..) + \alpha_{4j}(\text{READCAT}_{ij} - \overline{\text{READCAT}}..) \\ & + \alpha_{5j}(\text{GRDLVL}_{ij} - \overline{\text{GRDLVL}}..) + \alpha_{6j}(\text{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$

Table A4. 4. Fit Indices for the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3 Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	16909.0	16927.0	16927.7

Table A4. 5. Estimated Fixed Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3 Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	791.6900	13.1788	60.07	<.0001	--	--
Lexile0	α_{10}	0.5258	0.0392	13.41	<.0001	0.14	0.00
Age	α_{20}	-11.7732	3.8697	-3.04	0.0024	0.01	-0.04
MathCAT	α_{30}	7.7617	3.3231	2.34	0.0197	0.00	0.03
ReadCAT	α_{40}	27.7896	3.3569	8.28	<.0001	0.06	0.10
Grade Level	α_{50}	12.8552	4.4864	2.87	0.0042	0.01	0.05
TRTGroup	α_{60}	59.1368	12.2740	4.82	<.0001	0.02	0.21

Table A4. 6. Estimated Random Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3 Across Five Years of Data

Variance Component (Final Model)	Estimate	SE	*z*-value	*p*-value
σ^2	46068.00	1853.10	24.86	<.0001
τ_{00}	541.62	544.48	0.99	0.1599
Variance Component (Unconditional Model)				
σ^2	73593.00	2958.43	24.88	<.0001
τ_{00}	381.10	491.85	0.77	0.2192

Section 3: Full Hierarchical Linear Model for Cross-Sectional ITT Analysis of ReadCAT_1Year Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\begin{aligned}\text{READCAT\_1Y}_{ij} = {} & \alpha_{0j} + \alpha_{1j}(\text{READCAT0}_{ij} - \overline{\text{READCAT0}}..) + \alpha_{2j}(\text{WHITE}_{ij} - \overline{\text{WHITE}}..) \\ & + \alpha_{3j}(\text{AGE}_{ij} - \overline{\text{AGE}}..) + \alpha_{4j}(\text{MATHCAT}_{ij} - \overline{\text{MATHCAT}}..) \\ & + \alpha_{5j}(\text{LEXILE0}_{ij} - \overline{\text{LEXILE0}}..) + \alpha_{6j}(\text{DISB}_{ij} - \overline{\text{DISB}}..) \\ & + \alpha_{7j}(\text{GRDLVL}_{ij} - \overline{\text{GRDLVL}}..) + \alpha_{8j}(\text{MOBL}_{ij} - \overline{\text{MOBL}}..) \\ & + \alpha_{9j}(\text{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$

$$\alpha_{1j} = \alpha_{10}$$

$$\alpha_{2j} = \alpha_{20}$$

$$\alpha_{3j} = \alpha_{30}$$

$$\alpha_{4j} = \alpha_{40}$$

$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$
$$\alpha_{7j} = \alpha_{70}$$
$$\alpha_{8j} = \alpha_{80}$$
$$\alpha_{9j} = \alpha_{90}$$

Table A4. 7. Fit Indices for the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_1Year Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	983.8	1007.8	1007.1

Table A4. 8. Estimated Fixed Effects in the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_1Year Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	5.5646	0.2196	25.34	<.0001	--	--
ReadCAT0	α_{10}	0.4787	0.0737	6.49	<.0001	0.17	0.20
White	α_{20}	0.6542	0.3003	2.18	0.0305	0.02	0.28
Age	α_{30}	0.1095	0.0715	1.53	0.1274	0.01	0.05
MathCAT	α_{40}	0.1874	0.0669	2.80	0.0055	0.03	0.08
Lexile0	α_{50}	-0.0008	0.0008	-1.11	0.2686	0.01	-0.00
Disability	α_{60}	-0.3463	0.2627	-1.32	0.1888	0.01	-0.15
Grade Level	α_{70}	-0.0490	0.0925	-0.53	0.5967	0.00	-0.02
Mobility	α_{80}	-0.3605	0.2499	-1.44	0.1505	0.01	-0.15
TRTGroup	α_{90}	0.6131	0.2352	2.61	0.0097	0.03	0.26

Table A4. 9. Estimated Random Effects in the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_1Year Across Five Years of Data

Variance Component (Full Model)	Estimate	SE	*z*-value	*p*-value
σ^2	3.30	0.30	10.83	<.0001
τ_{00}	0.09	0.13	0.70	0.2430
Variance Component (Unconditional Model)				
σ^2	5.52	0.51	10.83	<.0001
τ_{00}	0.50	0.46	1.07	0.1419

Section 4: Final Hierarchical Linear Model for Cross-Sectional ITT Analysis of ReadCAT_1Year Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\begin{aligned}\text{READCAT\_1Y}_{ij} = {} & \alpha_{0j} + \alpha_{1j}(\text{READCAT0}_{ij} - \overline{\text{READCAT0}}..) + \alpha_{2j}(\text{WHITE}_{ij} - \overline{\text{WHITE}}..) \\ & + \alpha_{3j}(\text{MATHCAT}_{ij} - \overline{\text{MATHCAT}}..) + \alpha_{4j}(\text{DISB}_{ij} - \overline{\text{DISB}}..) \\ & + \alpha_{5j}(\text{MOBL}_{ij} - \overline{\text{MOBL}}..) + \alpha_{6j}(\text{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$

Table A4. 10. Fit Indices for the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_1Year Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	987.0	1005.0	1004.5

Table A4. 11. Estimated Fixed Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_1Year Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	5.5831	0.2357	23.68	<.0001	--	--
ReadCAT0	α_{10}	0.4417	0.0676	6.53	<.0001	0.18	0.19
White	α_{20}	0.7024	0.2978	2.36	0.0192	0.02	0.30
MathCAT	α_{30}	0.1816	0.0665	2.73	0.0068	0.03	0.08
Disability	α_{40}	-0.3403	0.2595	-1.31	0.1910	0.01	-0.14
Mobility	α_{50}	-0.3399	0.2446	-1.39	0.1659	0.01	-0.14
TRTGroup	α_{60}	0.6061	0.2358	2.57	0.0108	0.03	0.26

Table A4. 12. Estimated Random Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_1Year Across Five Years of Data

Variance Component (Final Model)	Estimate	SE	*z*-value	*p*-value
σ^2	3.33	0.31	10.82	<.0001
τ_{00}	0.13	0.16	0.79	0.2141
Variance Component (Unconditional Model)				
σ^2	5.52	0.51	10.83	<.0001
τ_{00}	0.50	0.46	1.07	0.1419

Section 5.A: Full Linear Regression Model for Cross-Sectional ITT Analysis of Lexile3_1YearCAT Across Five Years of Data

$$\begin{aligned}\text{LEXILE2}_i = {} & \alpha_0 + \alpha_1(\text{LEXILE0}_i - \overline{\text{LEXILE0.}}) + \alpha_2(\text{WHITE}_i - \overline{\text{WHITE.}}) + \alpha_3(\text{AGE}_i - \overline{\text{AGE.}}) \\ & + \alpha_4(\text{MATHCAT}_i - \overline{\text{MATHCAT.}}) + \alpha_5(\text{READCAT0}_i - \overline{\text{READCAT0.}}) + \alpha_6(\text{DISB}_i - \overline{\text{DISB.}}) \\ & + \alpha_7(\text{GRDLVL}_i - \overline{\text{GRDLVL.}}) + \alpha_8(\text{MOBL}_i - \overline{\text{MOBL.}}) + \alpha_9(\text{TRTGRP}_i) + \varepsilon_i\end{aligned}$$

Table A4. 13. Fit Indices for the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	3021.5	3043.5	3081.1

Table A4. 14: Estimated Regression Coefficients in the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

Predictors		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_0	745.7500	20.5375	36.31	<.0001	--	--
Lexile0	α_1	0.4928	0.0885	5.57	<.0001	0.14	0.00
White	α_2	-4.2296	33.3952	-0.13	0.8993	0.00	-0.02
Age	α_3	-6.6955	8.0921	-0.83	0.4089	0.00	-0.02
MathCAT	α_4	19.8111	7.4724	2.65	0.0086	0.03	0.07
ReadCAT0	α_5	15.8824	8.2758	1.92	0.0563	0.02	0.06
Disability	α_6	-42.1446	29.9980	-1.40	0.1615	0.01	-0.16
Grade Level	α_7	4.8973	10.5577	0.46	0.6432	0.00	0.02
Mobility	α_8	-26.6024	28.1610	-0.94	0.3459	0.00	-0.10
TRTGroup	α_9	74.0868	27.0892	2.73	0.0068	0.03	0.27

Table A4. 15. Estimated Error Variance in the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

Error Variance	Estimate	SE	*z*-value	*p*-value
σ^2	39784.00	3750.88	10.61	<.0001

Section 5.B: Full Hierarchical Linear Model for Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\begin{aligned}\text{LEXILE2}_{ij} = {} & \alpha_{0j} + \alpha_{1j}(\text{LEXILE0}_{ij} - \overline{\text{LEXILE0}}..) + \alpha_{2j}(\text{WHITE}_{ij} - \overline{\text{WHITE}}..) + \alpha_{3j}(\text{AGE}_{ij} - \overline{\text{AGE}}..) \\ & + \alpha_{4j}(\text{MATHCAT}_{ij} - \overline{\text{MATHCAT}}..) + \alpha_{5j}(\text{READCAT0}_{ij} - \overline{\text{READCAT0}}..) + \alpha_{6j}(\text{DISB}_{ij} - \overline{\text{DISB}}..) \\ & + \alpha_{7j}(\text{GRDLVL}_{ij} - \overline{\text{GRDLVL}}..) + \alpha_{8j}(\text{MOBL}_{ij} - \overline{\text{MOBL}}..) + \alpha_{9j}(\text{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$
$$\alpha_{7j} = \alpha_{70}$$
$$\alpha_{8j} = \alpha_{80}$$
$$\alpha_{9j} = \alpha_{90}$$

Table A4. 16. Fit Indices for the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	3021.5	3043.5	3043.0

Table A4. 17. Estimated Fixed Effects in the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	745.7500	20.5375	36.31	<.0001	--	--
Lexile0	α_{10}	0.4928	0.0885	5.57	<.0001	0.14	0.00
White	α_{20}	-4.2296	33.3952	-0.13	0.8993	0.00	-0.02
Age	α_{30}	-6.6955	8.0921	-0.83	0.4089	0.00	-0.02
MathCAT	α_{40}	19.8111	7.4724	2.65	0.0086	0.03	0.07
ReadCAT0	α_{50}	15.8824	8.2758	1.92	0.0562	0.02	0.06
Disability	α_{60}	-42.1446	29.9980	-1.40	0.1614	0.01	-0.16
Grade Level	α_{70}	4.8973	10.5577	0.46	0.6432	0.00	0.02
Mobility	α_{80}	-26.6024	28.1610	-0.94	0.3458	0.00	-0.10
TRTGroup	α_{90}	74.0868	27.0892	2.73	0.0067	0.03	0.27

Table A4 18. Estimated Random Effects in the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

Variance Component	Estimate	SE	*z*-value	*p*-value
σ^2	39784.00	3750.88	10.61	<.0001
τ_{00}	0.00	.	.	.

Section 6.A: Final Linear Regression Model for Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

$$\begin{aligned}\text{LEXILE2}_i = {} & \alpha_0 + \alpha_1(\text{LEXILE0}_i - \overline{\text{LEXILE0.}}) + \alpha_2(\text{MATHCAT}_i - \overline{\text{MATHCAT.}}) \\ & + \alpha_3(\text{READCAT0}_i - \overline{\text{READCAT0.}}) + \alpha_4(\text{DISB}_i - \overline{\text{DISB.}}) + \alpha_5(\text{TRTGRP}_i) + \varepsilon_i\end{aligned}$$

Table A4. 19. Fit Indices for the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	3023.2	3037.2	3061.1

Table A4. 20. Estimated Regression Coefficients in the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_0	745.2900	20.5977	36.18	<.0001	--	--
Lexile0	α_1	0.4990	0.0835	5.97	<.0001	0.16	0.00
MathCAT	α_2	19.7073	7.4519	2.64	0.0088	0.03	0.07
ReadCAT0	α_3	15.9886	7.8180	2.05	0.0420	0.02	0.06
Disability	α_4	-38.0643	28.3193	-1.34	0.1803	0.01	-0.14
TRTGroup	α_5	74.8714	27.1542	2.76	0.0063	0.03	0.28

Table A4. 21. Estimated Error Variance in the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

Error Variance	Estimate	SE	*z*-value	*p*-value
σ^2	40076.00	3778.44	10.61	<.0001

Section 6.B: Final Hierarchical Linear Model for Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\text{LEXILE2}_{ij} = \alpha_{0j} + \alpha_{1j}(\text{LEXILE0}_{ij} - \overline{\text{LEXILE0}}..) + \alpha_{2j}(\text{MATHCAT}_{ij} - \overline{\text{MATHCAT}}..)$$
$$+ \alpha_{3j}(\text{READCAT0}_{ij} - \overline{\text{READCAT0}}..) + \alpha_{4j}(\text{DISB}_{ij} - \overline{\text{DISB}}..) + \alpha_{5j}(\text{TRTGRP}_{ij}) + \varepsilon_{ij}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$

Table A4. 22. Fit Indices for the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	3023.2	3037.2	3036.8

Table A4. 23. Estimated Fixed Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	745.2900	20.5977	36.18	<.0001	--	--
Lexile0	α_{10}	0.4990	0.0835	5.97	<.0001	0.16	0.00
MathCAT	α_{20}	19.7073	7.4519	2.64	0.0088	0.03	0.07
ReadCAT0	α_{30}	15.9886	7.8180	2.05	0.0420	0.02	0.06
Disability	α_{40}	-38.0643	28.3193	-1.34	0.1803	0.01	-0.14
TRTGroup	α_{50}	74.8714	27.1542	2.76	0.0063	0.03	0.28

Table A4.24. Estimated Random Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_1YearCAT Across Five Years of Data

Variance Component	Estimate	SE	*z*-value	*p*-value
σ^2	40076.00	3778.44	10.61	<.0001
τ_{00}	0.00	.	.	.

Section 7.A: Full Linear Regression Model for Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

$$
\begin{aligned}
\text{READCAT\_LAST}_i = {} & \alpha_0 + \alpha_1(\text{READCAT0}_i - \overline{\text{READCAT0.}}) + \alpha_2(\text{WHITE}_i - \overline{\text{WHITE.}}) \\
& + \alpha_3(\text{AGE}_i - \overline{\text{AGE.}}) + \alpha_4(\text{MATHCAT}_i - \overline{\text{MATHCAT.}}) \\
& + \alpha_5(\text{LEXILE0}_i - \overline{\text{LEXILE0.}}) + \alpha_6(\text{DISB}_i - \overline{\text{DISB.}}) \\
& + \alpha_7(\text{GRDLVL}_i - \overline{\text{GRDLVL.}}) + \alpha_8(\text{MOBL}_i - \overline{\text{MOBL.}}) \\
& + \alpha_9(\text{TRTGRP}_i) + \varepsilon_i
\end{aligned}
$$

Table A4. 25. Fit Indices for the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	4240.4	4262.4	4315.6

Table A4. 26. Estimated Regression Coefficients in the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

Predictors		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_0	6.4355	0.1132	56.86	<.0001	--	--
ReadCAT0	α_1	0.3864	0.0450	8.58	<.0001	0.08	0.14
White	α_2	0.6177	0.1934	3.19	0.0015	0.01	0.23
Age	α_3	0.1157	0.0479	2.42	0.0159	0.01	0.04
MathCAT	α_4	0.1442	0.0439	3.29	0.0011	0.01	0.05
Lexile0	α_5	0.0015	0.0005	2.91	0.0037	0.01	0.00
Disability	α_6	-0.1755	0.1702	-1.03	0.3029	0.00	-0.06
Grade Level	α_7	-0.0024	0.0583	-0.04	0.9667	0.00	-0.00
Mobility	α_8	0.1493	0.1575	0.95	0.3432	0.00	0.05
TRTGroup	α_9	0.2623	0.1543	1.70	0.0896	0.00	0.10

Table A4. 27. Estimated Error Variance in the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

Error Variance	Estimate	SE	*z*-value	*p*-value
σ^2	5.49	0.25	21.61	<.0001

Section 7.B: Full Hierarchical Linear Model for Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\begin{aligned}\text{READCAT\_LAST}_{ij} = {} & \alpha_{0j} + \alpha_{1j}(\text{READCAT0}_{ij} - \overline{\text{READCAT0}}..) + \alpha_{2j}(\text{WHITE}_{ij} - \overline{\text{WHITE}}..) \\ & + \alpha_{3j}(\text{AGE}_{ij} - \overline{\text{AGE}}..) + \alpha_{4j}(\text{MATHCAT}_{ij} - \overline{\text{MATHCAT}}..) \\ & + \alpha_{5j}(\text{LEXILE0}_{ij} - \overline{\text{LEXILE0}}..) + \alpha_{6j}(\text{DISB}_{ij} - \overline{\text{DISB}}..) \\ & + \alpha_{7j}(\text{GRDLVL}_{ij} - \overline{\text{GRDLVL}}..) + \alpha_{8j}(\text{MOBL}_{ij} - \overline{\text{MOBL}}..) \\ & + \alpha_{9j}(\text{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$
$$\alpha_{7j} = \alpha_{70}$$
$$\alpha_{8j} = \alpha_{80}$$
$$\alpha_{9j} = \alpha_{90}$$

Table A4. 28. Fit Indices for the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	4240.4	4262.4	4261.8

Table A4. 29. Estimated Fixed Effects in the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	6.4355	0.1132	56.86	<.0001	--	--
ReadCAT0	α_{10}	0.3864	0.0450	8.58	<.0001	0.08	0.14
White	α_{20}	0.6177	0.1934	3.19	0.0015	0.01	0.23
Age	α_{30}	0.1157	0.0479	2.42	0.0159	0.01	0.04
MathCAT	α_{40}	0.1442	0.0439	3.29	0.0010	0.01	0.05
Lexile0	α_{50}	0.0015	0.0005	2.91	0.0037	0.01	0.00
Disability	α_{60}	-0.1755	0.1702	-1.03	0.3029	0.00	-0.06
Grade Level	α_{70}	-0.0024	0.0583	-0.04	0.9667	0.00	-0.00
Mobility	α_{80}	0.1493	0.1575	0.95	0.3432	0.00	0.05
TRTGroup	α_{90}	0.2623	0.1543	1.70	0.0896	0.00	0.10

Table A4. 30. Estimated Random Effects in the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

Variance Component	Estimate	SE	*z*-value	*p*-value
σ^2	5.49	0.25	21.61	<.0001
τ_{00}	0.00	.	.	.

Section 8.A: Final Linear Regression Model for Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

$$\begin{aligned}\text{READCAT\_LAST}_i = {} & \alpha_0 + \alpha_1(\text{READCAT0}_i - \overline{\text{READCAT0.}}) + \alpha_2(\text{WHITE}_i - \overline{\text{WHITE.}}) \\ & + \alpha_3(\text{AGE}_i - \overline{\text{AGE.}}) + \alpha_4(\text{MATHCAT}_i - \overline{\text{MATHCAT.}}) \\ & + \alpha_5(\text{LEXILE0}_i - \overline{\text{LEXILE0.}}) + \alpha_6(\text{TRTGRP}_i) + \varepsilon_i\end{aligned}$$

Table A4. 31. Fit Indices for the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	4242.4	4258.4	4297.2

Table A4. 32. Estimated Regression Coefficients in the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

Predictors		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_0	6.4423	0.1132	56.93	<.0001	--	--
ReadCAT0	α_1	0.3933	0.0445	8.84	<.0001	0.08	0.14
White	α_2	0.5653	0.1894	2.99	0.0029	0.01	0.21
Age	α_3	0.1118	0.0461	2.43	0.0154	0.01	0.04
MathCAT	α_4	0.1523	0.0427	3.57	0.0004	0.01	0.06
Lexile0	α_5	0.0015	0.0005	3.02	0.0026	0.01	0.00
TRTGroup	α_6	0.2497	0.1541	1.62	0.1055	0.00	0.09

Table A4. 33. Estimated Error Variance in the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

Error Variance	Estimate	SE	*z*-value	*p*-value
σ^2	5.50	0.25	21.61	<.0001

Section 8.B: Final Hierarchical Linear Model for Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\begin{aligned}\mathrm{READCAT\_LAST}_{ij} = {} & \alpha_{0j} + \alpha_{1j}(\mathrm{READCAT0}_{ij} - \overline{\mathrm{READCAT0}..}) + \alpha_{2j}(\mathrm{WHITE}_{ij} - \overline{\mathrm{WHITE}}..) \\ & + \alpha_{3j}(\mathrm{AGE}_{ij} - \overline{\mathrm{AGE}}..) + \alpha_{4j}(\mathrm{MATHCAT}_{ij} - \overline{\mathrm{MATHCAT}}..) \\ & + \alpha_{5j}(\mathrm{LEXILE0}_{ij} - \overline{\mathrm{LEXILE0}}..) + \alpha_{6j}(\mathrm{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$

Table A4. 34. Fit Indices for the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	4242.4	4258.4	4258.0

Table A4. 35. Estimated Fixed Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	6.4423	0.1132	56.93	<.0001	--	--
ReadCAT0	α_{10}	0.3933	0.0445	8.84	<.0001	0.08	0.14
White	α_{20}	0.5653	0.1894	2.99	0.0029	0.01	0.21
Age	α_{30}	0.1118	0.0461	2.43	0.0154	0.01	0.04
MathCAT	α_{40}	0.1523	0.0427	3.57	0.0004	0.01	0.06
Lexile0	α_{50}	0.0015	0.0005	3.02	0.0026	0.01	0.00
TRTGroup	α_{60}	0.2497	0.1541	1.62	0.1055	0.00	0.09

Table A4. 36. Estimated Random Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of ReadCAT_last Across Five Years of Data

Variance Component	Estimate	SE	*z*-value	*p*-value
σ^2	5.50	0.25	21.61	<.0001
τ_{00}	0.00	.	.	.

Section 9.A: Full Linear Regression Model for Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

$$\begin{aligned}\text{LEXILE2}_i = {} & \alpha_0 + \alpha_1(\text{LEXILE0}_i - \overline{\text{LEXILE0.}}) + \alpha_2(\text{WHITE}_i - \overline{\text{WHITE.}}) + \alpha_3(\text{AGE}_i - \overline{\text{AGE.}}) \\ & + \alpha_4(\text{MATHCAT}_i - \overline{\text{MATHCAT.}}) + \alpha_5(\text{READCAT0}_i - \overline{\text{READCAT0.}}) + \alpha_6(\text{DISB}_i - \overline{\text{DISB.}}) \\ & + \alpha_7(\text{GRDLVL}_i - \overline{\text{GRDLVL.}}) + \alpha_8(\text{MOBL}_i - \overline{\text{MOBL.}}) + \alpha_9(\text{TRTGRP}_i) + \varepsilon_i\end{aligned}$$

Table A4. 37. Fit Indices for the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	11718.4	11740.4	11792.9

Table A4. 38. Estimated Regression Coefficients in the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

Predictors		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_0	770.8400	10.5820	72.84	<.0001	--	--
Lexile0	α_1	0.5337	0.0468	11.41	<.0001	0.15	0.00
White	α_2	3.9610	17.8664	0.22	0.8246	0.00	0.01
Age	α_3	-11.0939	4.4418	-2.50	0.0127	0.01	-0.04
MathCAT	α_4	9.9501	4.0630	2.45	0.0145	0.01	0.04
ReadCAT0	α_5	27.1902	4.1429	6.56	<.0001	0.05	0.10
Disability	α_6	-18.0090	15.6883	-1.15	0.2513	0.00	-0.07
Grade Level	α_7	12.4932	5.3781	2.32	0.0204	0.01	0.05
Mobility	α_8	-0.1447	14.5646	-0.01	0.9921	0.00	-0.00
TRTGroup	α_9	70.2006	14.2729	4.92	<.0001	0.03	0.26

Table A4. 39. Estimated Error Variance in the Full Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

Error Variance	Estimate	SE	*z*-value	*p*-value
σ^2	43399.00	2084.43	20.82	<.0001

Section 9.B: Full Hierarchical Linear Model for Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\begin{aligned}\text{LEXILE2}_{ij} &= \alpha_{0j} + \alpha_{1j}(\text{LEXILE0}_{ij} - \overline{\text{LEXILE0}}..) + \alpha_{2j}(\text{WHITE}_{ij} - \overline{\text{WHITE}}..) + \alpha_{3j}(\text{AGE}_{ij} - \overline{\text{AGE}}..) \\ &+ \alpha_{4j}(\text{MATHCAT}_{ij} - \overline{\text{MATHCAT}}..) + \alpha_{5j}(\text{READCAT0}_{ij} - \overline{\text{READCAT0}}..) + \alpha_{6j}(\text{DISB}_{ij} - \overline{\text{DISB}}..) \\ &+ \alpha_{7j}(\text{GRDLVL}_{ij} - \overline{\text{GRDLVL}}..) + \alpha_{8j}(\text{MOBL}_{ij} - \overline{\text{MOBL}}..) + \alpha_{9j}(\text{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$
$$\alpha_{7j} = \alpha_{70}$$
$$\alpha_{8j} = \alpha_{80}$$
$$\alpha_{9j} = \alpha_{90}$$

Table A4. 40. Fit Indices for the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	11718.4	11740.4	11739.8

Table A4. 41. Estimated Fixed Effects in the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	770.8400	10.5820	72.84	<.0001	--	--
Lexile0	α_{10}	0.5337	0.0468	11.41	<.0001	0.15	0.00
White	α_{20}	3.9610	17.8664	0.22	0.8246	0.00	0.01
Age	α_{30}	-11.0939	4.4418	-2.50	0.0127	0.01	-0.04
MathCAT	α_{40}	9.9501	4.0630	2.45	0.0145	0.01	0.04
ReadCAT0	α_{50}	27.1902	4.1429	6.56	<.0001	0.05	0.10
Disability	α_{60}	-18.0090	15.6883	-1.15	0.2513	0.00	-0.07
Grade Level	α_{70}	12.4932	5.3781	2.32	0.0204	0.01	0.05
Mobility	α_{80}	-0.1447	14.5646	-0.01	0.9921	0.00	-0.00
TRTGroup	α_{90}	70.2006	14.2729	4.92	<.0001	0.03	0.26

Table A4. 42. Estimated Random Effects in the Full Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

Variance Component	Estimate	SE	*z*-value	*p*-value
σ^2	43399.00	2084.43	20.82	<.0001
τ_{00}	0.00	.	.	.

Section 10.A: Final Linear Regression Model for Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

$$\begin{aligned}\mathrm{LEXILE2}_i = \alpha_0 + \alpha_1(\mathrm{LEXILE0}_i - \overline{\mathrm{LEXILE0.}}) + \alpha_2(\mathrm{AGE}_i - \overline{\mathrm{AGE.}}) + \alpha_3(\mathrm{MATHCAT}_i - \overline{\mathrm{MATHCAT.}}) \\ + \alpha_4(\mathrm{READCAT0}_i - \overline{\mathrm{READCAT0.}}) + \alpha_5(\mathrm{GRDLVL}_i - \overline{\mathrm{GRDLVL.}}) + \alpha_6(\mathrm{TRTGRP}_i) + \varepsilon_i\end{aligned}$$

Table A4. 43. Fit Indices for the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	11719.8	11735.8	11773.9

Table A4. 44. Estimated Regression Coefficients in the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

Predictors		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_0	771.3900	10.5776	72.93	<.0001	--	--
Lexile0	α_1	0.5394	0.0463	11.64	<.0001	0.16	0.00
Age	α_2	-11.1494	4.4338	-2.51	0.0121	0.01	-0.04
MathCAT	α_3	10.9252	3.9738	2.75	0.0061	0.01	0.04
ReadCAT0	α_4	27.7827	4.0277	6.90	<.0001	0.06	0.10
Grade Level	α_5	11.9186	5.3317	2.24	0.0256	0.01	0.04
TRTGroup	α_6	69.1990	14.2534	4.85	<.0001	0.03	0.25

Table A4. 45. Estimated Error Variance in the Final Cross-Sectional Regression Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

Error Variance	Estimate	SE	*z*-value	*p*-value
σ^2	43465.00	2087.61	20.82	<.0001

Section 10.B: Final Hierarchical Linear Model for Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

For student *i* in institution *j*,

Level 1:

$$\begin{aligned}\mathrm{LEXILE2}_{ij} &= \alpha_{0j} + \alpha_{1j}(\mathrm{LEXILE0}_{ij} - \overline{\mathrm{LEXILE0..}}) + \alpha_{2j}(\mathrm{AGE}_{ij} - \overline{\mathrm{AGE..}}) \\ &+ \alpha_{3j}(\mathrm{MATHCAT}_{ij} - \overline{\mathrm{MATHCAT..}}) + \alpha_{4j}(\mathrm{READCAT0}_{ij} - \overline{\mathrm{READCAT0..}}) \\ &+ \alpha_{5j}(\mathrm{GRDLVL}_{ij} - \overline{\mathrm{GRDLVL..}}) + \alpha_{6j}(\mathrm{TRTGRP}_{ij}) + \varepsilon_{ij}\end{aligned}$$

Level 2:

$$\alpha_{0j} = \alpha_{00} + u_{0j}$$
$$\alpha_{1j} = \alpha_{10}$$
$$\alpha_{2j} = \alpha_{20}$$
$$\alpha_{3j} = \alpha_{30}$$
$$\alpha_{4j} = \alpha_{40}$$
$$\alpha_{5j} = \alpha_{50}$$
$$\alpha_{6j} = \alpha_{60}$$

Table A4. 46. Fit Indices for the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	11719.8	11735.8	11735.3

Table A4. 47. Estimated Fixed Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2	Glass's Δ
Intercept	α_{00}	771.3900	10.5776	72.93	<.0001	--	--
Lexile0	α_{10}	0.5394	0.0463	11.64	<.0001	0.16	0.00
Age	α_{20}	-11.1494	4.4338	-2.51	0.0121	0.01	-0.04
MathCAT	α_{30}	10.9252	3.9738	2.75	0.0061	0.01	0.04
ReadCAT0	α_{40}	27.7827	4.0277	6.90	<.0001	0.05	0.10
Grade Level	α_{50}	11.9186	5.3317	2.24	0.0256	0.01	0.04
TRTGroup	α_{60}	69.1990	14.2534	4.85	<.0001	0.03	0.25

Table A4. 48. Estimated Random Effects in the Final Cross-Sectional HLM Model: Cross-Sectional ITT Analysis of SRI Lexile3_lastCAT Across Five Years of Data

Variance Component	Estimate	SE	*z*-value	*p*-value
σ^2	43465.00	2087.61	20.82	<.0001
τ_{00}	0.00	.	.	.

Appendix A5: Additional Analysis: Longitudinal SRI HLM Descriptive Statistics and Estimates Across Five Years of Data

Section 11: Descriptive Statistics for the Hierarchical Linear Model for Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

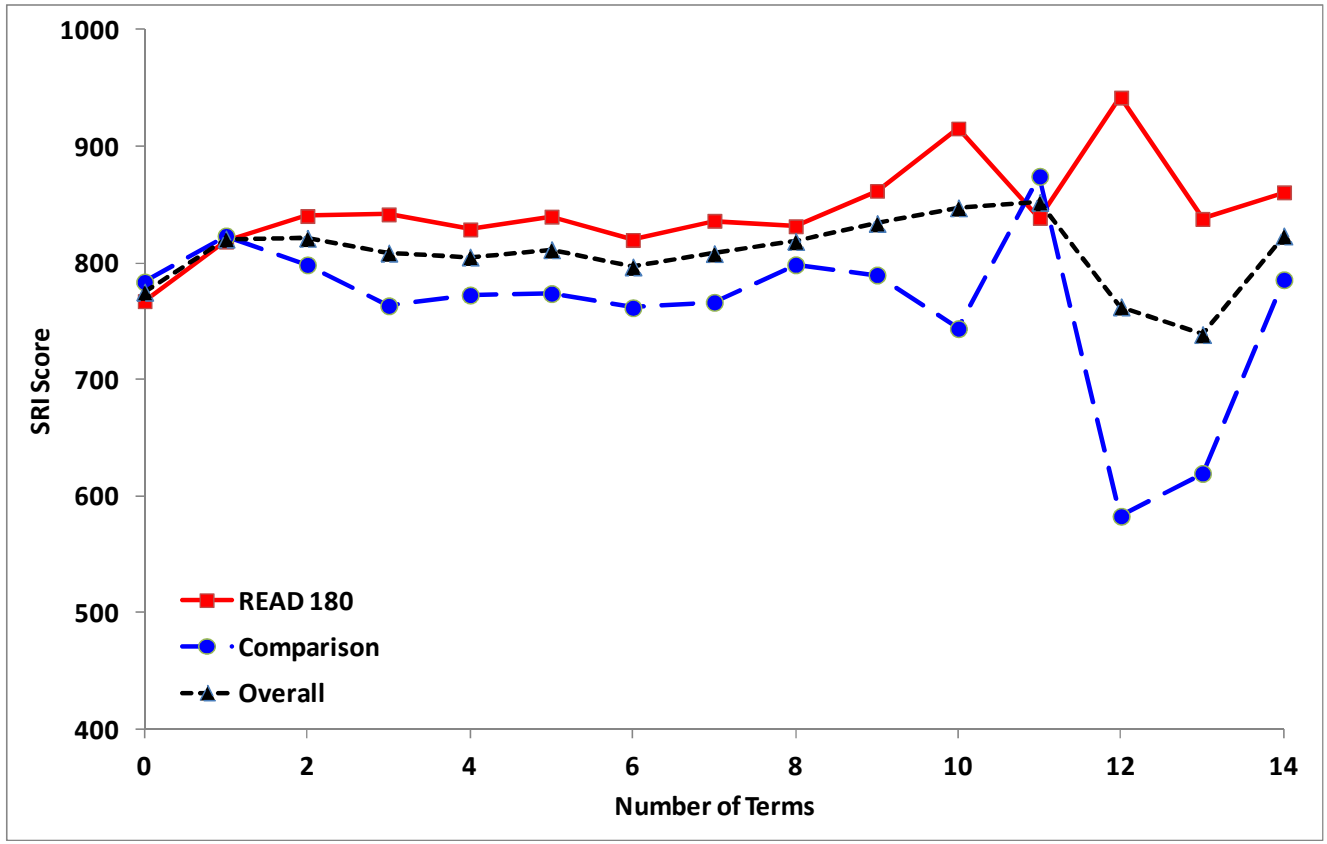


Figure A5. 1. Time Plot of the Mean Responses for the READ 180 Group and the Comparison Group.

Table A5. 1. Mean SRI Lexile Scores at Different Measurement Occasions for the READ 180 Group, the Comparison Group, and the Overall Across Five Years of Data

	SRI_0	SRI_1	SRI_2	SRI_3	SRI_4	SRI_5	SRI_6	SRI_7	SRI_8	SRI_9	SRI_{10}	SRI_{11}	SRI_{12}	SRI_{13}	SRI_{14}
READ 180	767.33	818.52	840.38	841.84	828.94	839.84	819.90	836.30	831.59	861.97	915.50	838.86	942.00	837.83	860.75
Comparison	783.82	823.34	798.52	763.40	772.62	773.85	761.73	766.39	798.58	789.79	743.80	874.46	582.89	619.60	785.75
Overall	775.05	820.69	821.28	808.61	805.15	811.42	796.72	808.12	818.58	833.98	847.37	852.09	762.44	738.64	823.25

Table A5. 2. Standard Deviations of SRI Lexile Scores at Different Measurement Occasions for the READ 180 Group, the Comparison Group, and the Overall Across Five Years of Data

	SRI_0	SRI_1	SRI_2	SRI_3	SRI_4	SRI_5	SRI_6	SRI_7	SRI_8	SRI_9	SRI_{10}	SRI_{11}	SRI_{12}	SRI_{13}	SRI_{14}
READ 180	193.65	263.18	264.48	266.16	267.28	254.20	269.03	277.95	302.62	302.75	272.91	343.38	336.76	359.59	149.65
Comparison	189.90	265.81	280.96	303.24	303.85	307.19	309.19	304.70	292.88	257.75	336.40	347.89	439.12	543.01	469.27
Overall	192.01	264.27	272.81	284.97	284.48	279.91	286.74	290.49	298.38	286.95	308.91	340.36	422.19	442.25	324.94

Table A5. 3. Number of Youth at Different Measurement Occasions for the READ 180 Group, the Comparison Group, and the Overall Across Five Years of Data

	SRI_0	SRI_1	SRI_2	SRI_3	SRI_4	SRI_5	SRI_6	SRI_7	SRI_8	SRI_9	SRI_{10}	SRI_{11}	SRI_{12}	SRI_{13}	SRI_{14}
READ 180	741	724	677	589	447	304	231	154	103	60	38	22	9	6	4
Comparison	652	593	568	433	327	230	153	104	67	38	25	13	9	5	4
Overall	1393	1317	1245	1022	774	534	384	258	170	98	63	35	18	11	8

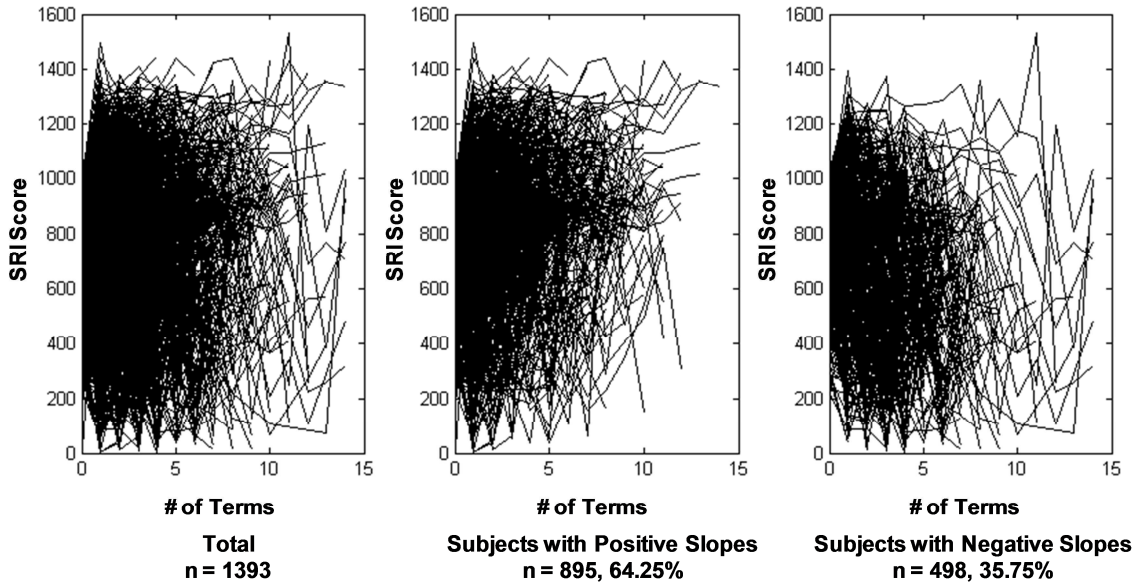


*Figure A5. 2. Spaghetti plots for **the Overall Group** in Total (Left Panel), Subjects with Positive Slopes (Middle Panel), and Subjects with Negative or Zero Slopes (Right Panel).*

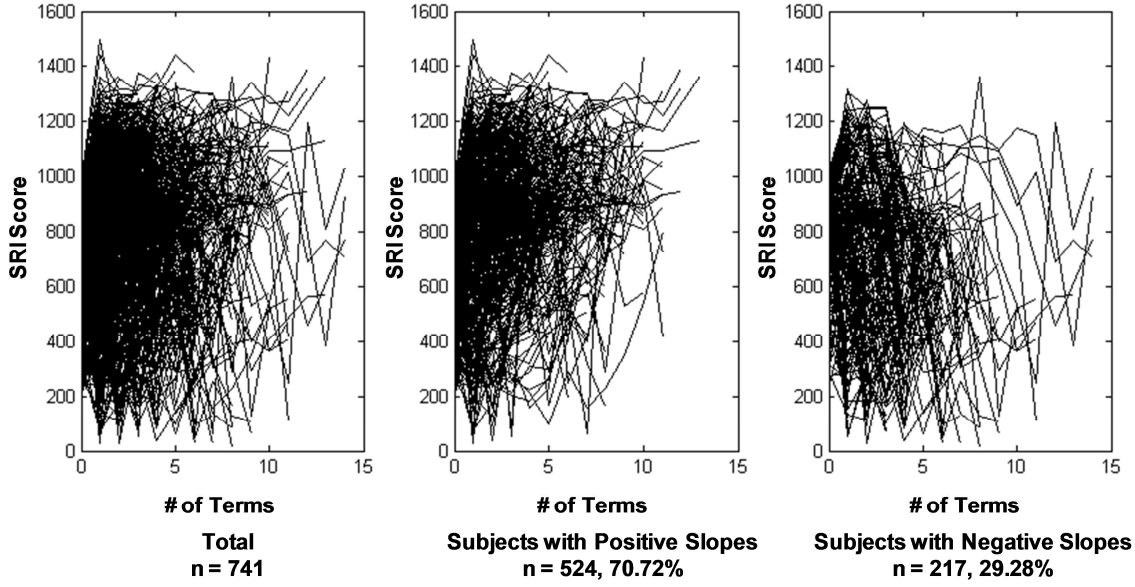


*Figure A5. 3. Spaghetti plots for **the READ 180 Group** in Total (Left Panel), Subjects with Positive Slopes (Middle Panel), and Subjects with Negative or Zero Slopes (Right Panel).*

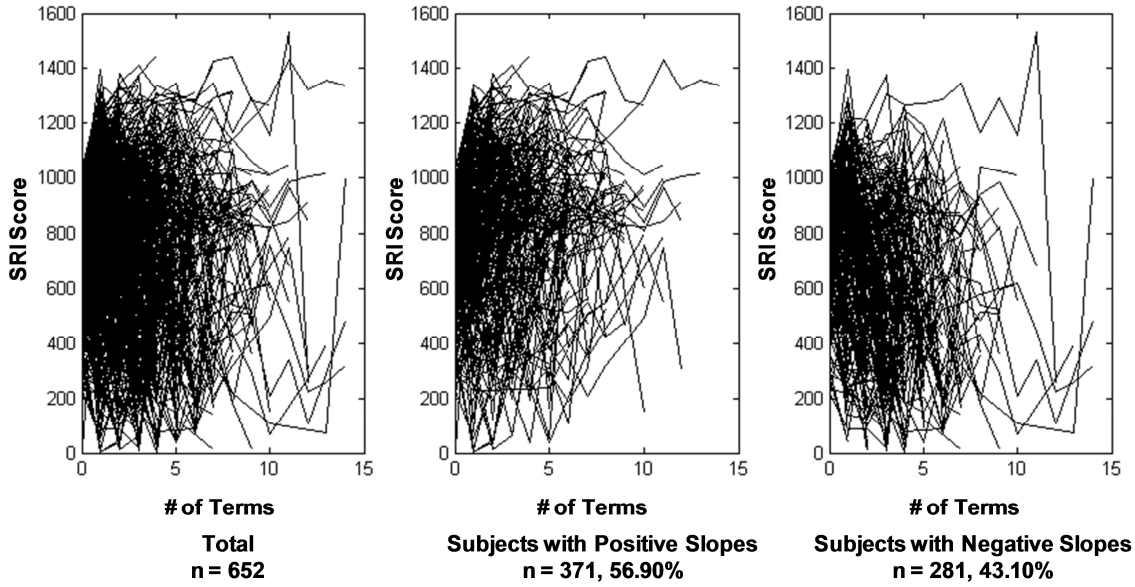


*Figure A5. 4. Spaghetti plots for **the Comparison Group** in Total (Left Panel), Subjects with Positive Slopes (Middle Panel), and Subjects with Negative or Zero Slopes (Right Panel).*

Table A5. 4. Number and Percentage of Subjects with Positive or Negative Growth Slopes in the READ 180 Group, the Comparison Group, and the Overall

	Overall		READ 180		Comparison	
	n	col %	n	col %	n	col %
Slope > 0	895	64.25%	524	70.72%	371	56.90%
Slope ≤ 0	498	35.75%	217	29.28%	281	43.10%
Total	1393	100%	741	100%	652	100%

Section 12: Full Hierarchical Linear Model for Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

Level 1:

$$y_{ij} = \alpha_i + j\,\beta_i + \varepsilon_{ij}\text{ , for } i = 1,2,...,n \text{ and } j = 0,1,2,...N_i$$

Level 2:

$$\alpha_i = \alpha_0 + \alpha_1(\text{WHITE}_i - \overline{\text{WHITE.}}) + \alpha_2(\text{AGE}_i - \overline{\text{AGE.}}) + \alpha_3(\text{MATHCAT}_i - \overline{\text{MATHCAT.}})$$
$$+ \alpha_4(\text{READCAT}_i - \overline{\text{READCAT.}}) + \alpha_5(\text{DISB}_i - \overline{\text{DISB.}}) + \alpha_6(\text{GRDLVL}_i - \overline{\text{GRDLVL.}})$$
$$+ \alpha_7(\text{INST}_i - \overline{\text{INST.}}) + \alpha_8(\text{MOBL}_i - \overline{\text{MOBL.}}) + \alpha_9(\text{TRTGRP}_i) + b_{0i}$$

$$\beta_i = \beta_0 + \beta_1(\text{WHITE}_i - \overline{\text{WHITE.}}) + \beta_2(\text{AGE}_i - \overline{\text{AGE.}}) + \beta_3(\text{MATHCAT}_i - \overline{\text{MATHCAT.}})$$
$$+ \beta_4(\text{READCAT}_{ii} - \overline{\text{READCAT.}}) + \beta_5(\text{DISB}_i - \overline{\text{DISB.}}) + \beta_6(\text{GRDLVL}_i - \overline{\text{GRDLVL.}})$$
$$+ \beta_7(\text{INST}_i - \overline{\text{INST.}}) + \beta_8(\text{MOBL}_i - \overline{\text{MOBL.}}) + \beta_9(\text{TRTGRP}_i) + b_{1i}$$

Table A5. 5. Fit Indices for the Full Linear Model: Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Full Linear Model	98002.7	98074.7	98263.3

Table A5. 6. Estimated Fixed Effects in the Full Linear Model: Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2
Intercept	α_0	811.5700	25.2703	32.12	<.0001	--
White	α_1	-17.0162	11.9893	-1.42	0.1560	0.00
Age	α_2	6.3733	3.0670	2.08	0.0379	0.00
MathCAT	α_3	9.1974	2.5368	3.63	0.0003	0.01
ReadCAT	α_4	36.3991	2.4883	14.63	<.0001	0.15
Disability	α_5	-48.5962	10.4848	-4.63	<.0001	0.02
Grade Level	α_6	17.4399	3.4547	5.05	<.0001	0.02
Inst_1		-34.3668	28.3351	-1.21	0.2254	
Inst_2		-23.8288	26.3267	-0.91	0.3655	
Inst_3		-24.6591	37.3107	-0.66	0.5088	
Inst_4	α_7	-18.5606	26.5378	-0.70	0.4844	0.00
Inst_5		-13.4195	29.8190	-0.45	0.6528	
Inst_6		70.4661	135.2800	0.52	0.6025	
Inst_7		-29.8588	27.2589	-1.10	0.2735	
Mobility	α_8	-0.7428	9.7149	-0.08	0.9391	0.00
TRTGroup	α_9	3.0805	9.3660	0.33	0.7423	0.00
Time	β_0	14.5855	9.7648	1.49	0.1355	0.00
White*Time	β_1	7.3966	4.3154	1.71	0.0869	0.00
Age*Time	β_2	-2.6818	1.0896	-2.46	0.0140	0.01
MathCAT*Time	β_3	-0.3079	0.9118	-0.34	0.7357	0.00
ReadCAT*Time	β_4	2.6666	0.8850	3.01	0.0027	0.01
Disability*Time	β_5	2.3270	3.5909	0.65	0.5171	0.00
Grade Level*Time	β_6	-1.6178	1.2180	-1.33	0.1844	0.00
Inst_1*Time		-31.9394	10.6572	-3.00	0.0028	
Inst_2*Time		-9.7853	10.0774	-0.97	0.3317	
Inst_3*Time		0.7837	14.3580	0.05	0.9565	
Inst_4*Time	β_7	-19.2272	10.3312	-1.86	0.0630	0.03
Inst_5*Time		-5.3774	11.0929	-0.48	0.6279	
Inst_6*Time		7.9543	73.4140	0.11	0.9137	
Inst_7*Time		-9.7615	10.3496	-0.94	0.3458	
Mobility*Time	β_8	10.8510	3.4512	3.14	0.0017	0.01
TRTGroup*Time	β_9	18.3419	3.2355	5.67	<.0001	0.04

Table A5. 7. Estimated Random Effects in the Full Linear Model: Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

Random Effect	b_0	b_1
b_0	17027*	
b_1	-363.94	1352.59*
ε	24603*	

Note. * *p*-value < .05

Section 13: Final Hierarchical Linear Model for Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

Level 1:

$$y_{ij} = \alpha_i + j\beta_i + \varepsilon_{ij} \text{ , for } i = 1,2,...,n \text{ and } j = 0,1,2,...N_i$$

Level 2:

$$\alpha_i = \alpha_0 + \alpha_1(\text{WHITE}_i - \overline{\text{WHITE.}}) + \alpha_2(\text{AGE}_i - \overline{\text{AGE.}}) + \alpha_3(\text{MATHCAT}_i - \overline{\text{MATHCAT.}})$$
$$+ \alpha_4(\text{READCAT}_i - \overline{\text{READCAT.}}) + \alpha_5(\text{DISB}_i - \overline{\text{DISB.}}) + \alpha_6(\text{GRDLVL}_i - \overline{\text{GRDLVL.}})$$
$$+ \alpha_7(\text{MOBL}_i - \overline{\text{MOBL.}}) + \alpha_8(\text{TRTGRP}_i) + b_{0i}$$

$$\beta_i = \beta_0 + \beta_1(\text{WHITE}_i - \overline{\text{WHITE.}}) + \beta_2(\text{AGE}_i - \overline{\text{AGE.}}) + \beta_3(\text{READCAT}_{ii} - \overline{\text{READCAT.}})$$
$$+ \beta_4(\text{MOBL}_i - \overline{\text{MOBL.}}) + \beta_5(\text{TRTGRP}_i) + b_{1i}$$

Table A5. 8. Fit Indices for the Final Linear Model: Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

	-2 (log-likelihood)	AIC	BIC
Final Linear Model	98045.5	98083.5	98183.0

Table A5. 9. Estimated Fixed Effects in the Final Linear Model: Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

Fixed Effect		Estimate	SE	*t*-ratio	*p*-value	Cohen's f^2
Intercept	α_0	788.2600	6.8633	114.85	<.0001	--
White	α_1	-17.4280	11.5573	-1.51	0.1318	0.00
Age	α_2	5.6970	2.7758	2.05	0.0403	0.00
MathCAT	α_3	8.9570	2.3167	3.87	0.0001	0.01
ReadCAT	α_4	37.1565	2.3916	15.54	<.0001	0.15
Disability	α_5	-44.7786	9.5611	-4.68	<.0001	0.02
Grade Level	α_6	16.5711	3.1226	5.31	<.0001	0.02
Mobility	α_7	-2.8732	9.4186	-0.31	0.7604	0.00
TRTGroup	α_8	3.3997	9.3415	0.36	0.7160	0.00
Time	β_0	0.0973	2.4771	0.04	0.9687	0.00
White*Time	β_1	10.0024	4.0330	2.48	0.0133	0.01
Age*Time	β_2	-4.6223	0.9542	-4.84	<.0001	0.03
ReadCAT*Time	β_3	1.9004	0.7003	2.71	0.0068	0.01
Mobility*Time	β_4	9.6964	3.3535	2.89	0.0039	0.01
TRTGroup*Time	β_5	19.5591	3.2662	5.99	<.0001	0.04

Table A5. 10. Estimated Random Effects in the Final Linear Model: Longitudinal ITT Analysis of SRI Lexile Scores Across Five Years of Data

Random Effect	b_0	b_1
b_0	17038*	
b_1	-324.64	1430.74*
ε	24606*	

Note. * *p*-value < .05

Appendix A6: ReadCAT Supplemental Descriptive Analyses

Table A6. 1. Frequency and Percentage of Students with the Number of Quarters between Baseline and Outcome ReadCAT Tests for the ReadCAT_1Year Analysis Sample

# of Quarters between Baseline and Outcome ReadCAT Tests	Comparison		Read 180	
	Frequency	Column Percentage	Frequency	Column Percentage
More than 3 quarters but no more than 4 quarters	48	43.64%	57	42.86%
More than 4 quarters but no more than 5 quarters	62	56.36%	76	57.14%
Total	110	100.00%	133	100.00%

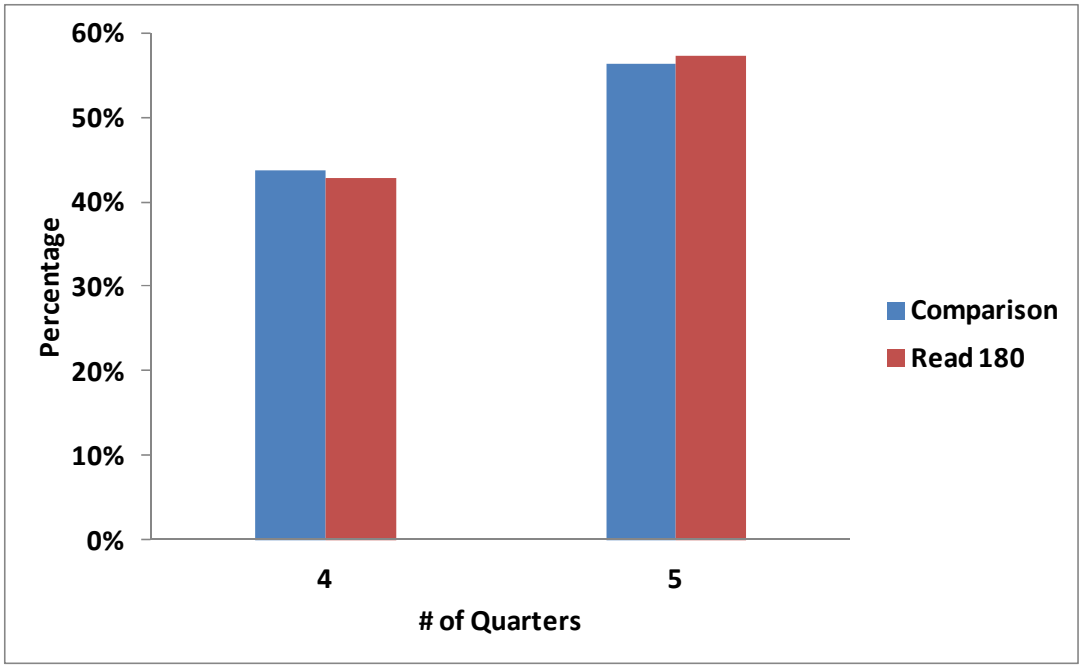


Figure A6. 1. Frequency Distribution of the Number of Quarters between Baseline and Outcome ReadCAT Tests for the ReadCAT_1Year Analysis Sample

Table A6. 2. Mean Outcome Scores and Mean Gain Scores for the ReadCAT_1Year Analysis Sample

# of Quarters between Baseline and Outcome ReadCAT Tests	Comparison		Read 180	
	Mean ReadCAT_1Year Score	Mean Gain Score	Mean ReadCAT_1Year Score	Mean Gain Score
More than 3 quarters but no more than 4 quarters	5.91	0.15	5.68	0.52
More than 4 quarters but no more than 5 quarters	5.41	-0.29	6.35	0.64

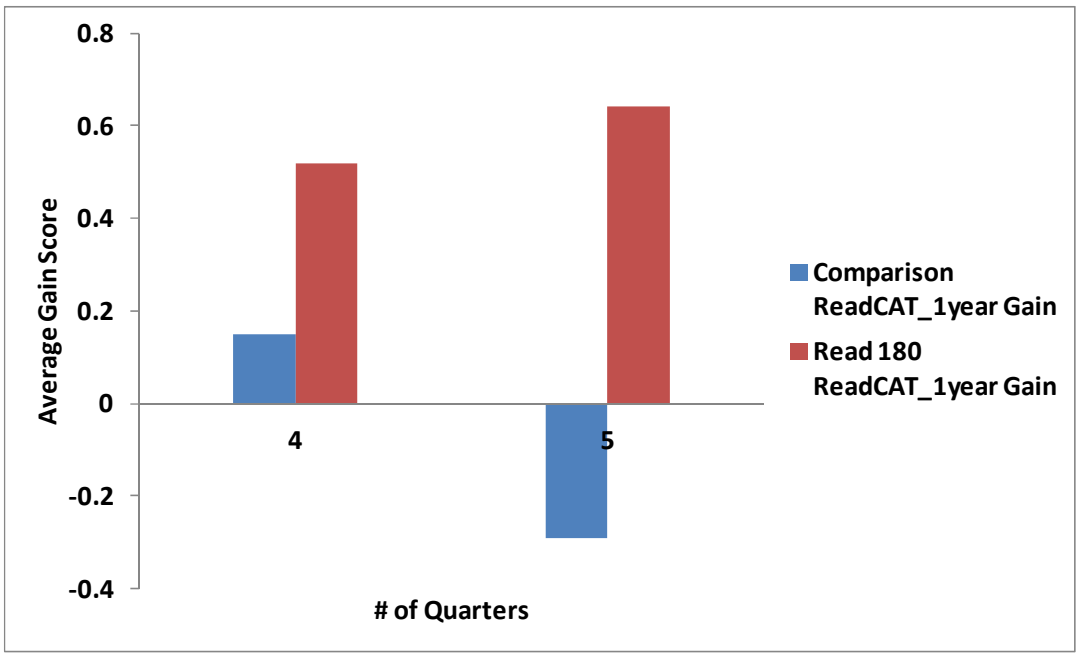


Figure A6. 2. Mean Gain Scores for the ReadCAT_1Year Analysis Sample by Treatment Group

Table A6. 3. Frequency and Percentage of Students with the Number of Quarters between Baseline and Outcome ReadCAT Tests for the ReadCAT_Last Analysis Sample

# of Quarters between Baseline and Outcome ReadCAT Tests	Comparison		Read 180	
	Frequency	Column Percentage	Frequency	Column Percentage
More than 0 day but no more than 1 quarter	14	3.26%	13	2.58%
More than 1 quarter but no more than 2 quarters	40	9.30%	46	9.13%
More than 2 quarters but no more than 3 quarters	59	13.72%	65	12.90%
More than 3 quarters but no more than 4 quarters	56	13.02%	71	14.09%
More than 4 quarters but no more than 5 quarters	56	13.02%	72	14.29%
More than 5 quarters but no more than 6 quarters	52	12.09%	55	10.91%
More than 6 quarters but no more than 7 quarters	39	9.07%	40	7.94%
More than 7 quarters but no more than 8 quarters	30	6.98%	34	6.75%
More than 8 quarters but no more than 9 quarters	26	6.05%	29	5.75%
More than 9 quarters but no more than 10 quarters	17	3.95%	27	5.36%
More than 10 quarters but no more than 11 quarters	13	3.02%	18	3.57%
More than 11 quarters but no more than 12 quarters	8	1.86%	10	1.98%
More than 12 quarters but no more than 13 quarters	11	2.56%	8	1.59%
More than 13 quarters but no more than 14 quarters	3	0.70%	5	0.99%
More than 14 quarters but no more than 15 quarters	4	0.93%	5	0.99%
More than 15 quarters but no more than 16 quarters	2	0.47%	6	1.19%
Total	430	100.00%	504	100.00%

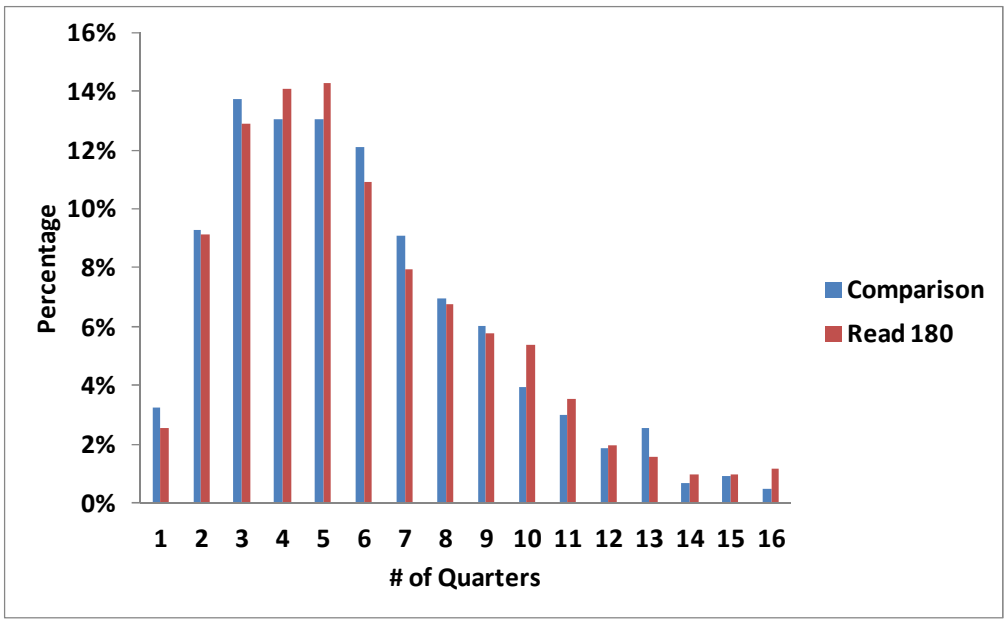


Figure A6. 3. Frequency Distribution of the Number of Quarters between Baseline and Outcome ReadCAT Tests for the ReadCAT_Last Analysis Sample

Table A6. 4. Mean Outcome Scores and Mean Gain Scores for the ReadCAT_Last Analysis Sample

# of Quarters between Baseline and Outcome ReadCAT Tests	Comparison		Read 180	
	Mean ReadCAT_ Last Score	Mean Gain Score	Mean ReadCAT_ Last Score	Mean Gain Score
More than 0 day but no more than 1 quarter	8.35	2.47	8.18	1.69
More than 1 quarter but no more than 2 quarters	6.75	1	6.35	0.08
More than 2 quarters but no more than 3 quarters	5.69	0.12	7.31	1.13
More than 3 quarters but no more than 4 quarters	6.26	0.23	6.28	0.75
More than 4 quarters but no more than 5 quarters	5.69	0.06	6.54	0.78
More than 5 quarters but no more than 6 quarters	6.72	1.2	6.75	0.81
More than 6 quarters but no more than 7 quarters	5.98	0.6	6	0.48
More than 7 quarters but no more than 8 quarters	6.78	1.23	6.18	0.88
More than 8 quarters but no more than 9 quarters	7.63	1.56	6.98	2.15
More than 9 quarters but no more than 10 quarters	5.57	-0.09	7.27	1.77
More than 10 quarters but no more than 11 quarters	7.88	0.95	6.16	1.43
More than 11 quarters but no more than 12 quarters	8.94	1.19	6.84	1.17
More than 12 quarters but no more than 13 quarters	6.19	1.56	8.04	1.81
More than 13 quarters but no more than 14 quarters	4.8	1.03	8.94	2.7
More than 14 quarters but no more than 15 quarters	7.72	2.63	7.3	3.32
More than 15 quarters but no more than 16 quarters	6.65	1	5.98	1.53

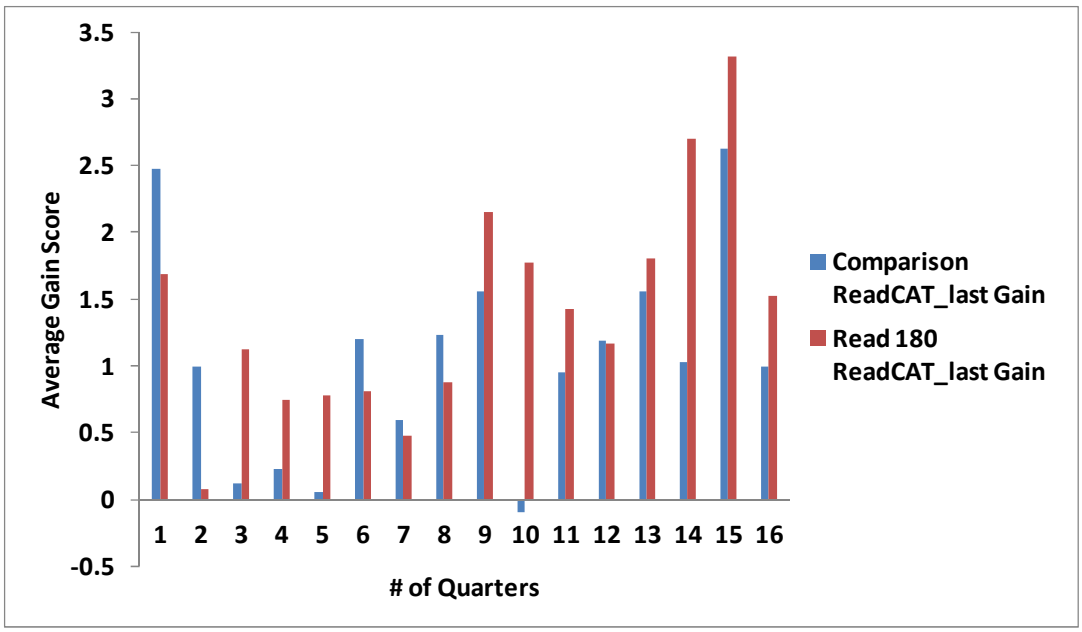


Figure A6. 4. Mean Gain Scores for the ReadCAT_Last Analysis Sample by Treatment Group

Table A6. 5. Students Included in Both ReadCAT_1Year and ReadCAT_Last Analysis Samples

# of Quarters between Baseline and Outcome ReadCAT Tests	Treatment Group	Included in ReadCAT_1Year Sample	Included in ReadCAT_Last Sample	Included in Both Samples
More than 3 quarters but no more than 4 quarters	Comparison	48	56	34
	Read 180	57	71	39
More than 4 quarters but no more than 5 quarters	Comparison	62	56	37
	Read 180	76	72	53

Appendix A7. Tests of Equivalency

In an effort to rule out competing interpretations and to help confirm the equivalence of the randomly assigned treatment groups at baseline across all five years of data, a series of analyses of variance (ANOVAs) were estimated. Two different dependent variables were used in the cross sectional analyses: the SRI and the ReadCAT (one year); the SRI was also used in the longitudinal analysis. Using the SRI and ReadCAT scores as dependent variables, two-way Anova was used to: 1. establish the equivalency of the randomly assigned youth to treatment condition at baseline; and 2. ensure that the students who were removed from the analysis were not statistically different from those who were kept in the analysis. The analyses in tables 7.1 and 7.2 focus on the cross-sectional HLM with the SRI variable as outcome. The results showed that those who were in the analysis were not significantly different (based on effect sizes) than those excluded from the HLM analysis and this difference did not depend on whether the youth was randomly assigned to the Read 180 or traditional English classroom. Further, there was no significant difference between the Read 180 and the Traditional groups with respect to the performance on the SRI at baseline.

Table 7.1. Descriptive Statistics by Treatment Groups and SRI Cross-Sectional HLM Analysis Status: Baseline SRI as Outcome

	HLM Analysis	Mean	SD	N
Read180	Out of the Analysis	767.37	207.29	381
	In the Analysis	770.00	189.94	677
	Total	769.05	196.27	1058
Traditional	Out of the Analysis	782.53	196.84	356
	In the Analysis	787.58	184.61	568
	Total	785.64	189.33	924
Total	Out of the Analysis	774.69	202.32	737
	In the Analysis	778.02	187.66	1245
	Total	776.78	193.19	1982

Table 7.2. SRI Cross-Sectional Analysis of Variance Source Table: Baseline SRI as Outcome

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Powerb
Corrected Model	142947.005a	3	47649.00	1.277	.28	.00	3.83	.34
Intercept	1.114E9	1	1.114E9	29848.83	.00	.93	29848.83	1.00
TRTGroup	123643.92	1	123643.92	3.31	.06	.00	3.31	.44
HLMstatus	6820.25	1	6820.25	.18	.66	.00	.18	.07
TRTGroupY * HLMstatus	679.93	1	679.93	.01	.89	.00	.01	.05
Error	73795304.271	1978	37308.041					
Total	1.270E9	1982						

a. R Squared = .002 (Adjusted R Squared = .000)
b. Computed using alpha = .05

Tables 7.3 and 7.4 present the comparable analysis using the ReadCAT baseline as the dependent variable. The results showed that those who were in the analysis were not significantly different (based on effect sizes) than those excluded from the HLM analysis and this difference did not depend on whether the youth was randomly assigned to the Read 180 or traditional English classroom. Further, there was no significant difference between the Read 180 and the Traditional groups with respect to the performance on the ReadCAT at baseline.

Table 7.3. Descriptive Statistics by Treatment Groups and SRI Cross-Sectional HLM Analysis Status: ReadCAT baseline as Outcome

	HLM Analysis	Mean	SD	N
Read180	Out of the Analysis	6.49	2.81	270
	In the Analysis	5.95	2.52	677
	Total	6.10	2.61	947
Traditional	Out of the Analysis	6.10	2.75	256
	In the Analysis	6.15	2.52	568
	Total	6.13	2.59	824
Total	Out of the Analysis	6.30	2.79	526
	In the Analysis	6.04	2.52	1245
	Total	6.12	2.60	1771

Table 7.4. SRI Cross-Sectional Analysis of Variance Source Table: ReadCAT baseline as Outcome

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Powerb
Corrected Model	56.93a	3	18.97	2.80	.03	.00	8.40	.67
Intercept	56270.57	1	56270.57	8306.859	.00	.82	8306.85	1.00
TRTGroupY1Y2Y3Y4Y5	3.43	1	3.43	.507	.47	.00	.50	.11
HLMstatusY5	22.41	1	22.41	3.309	.06	.00	3.30	.44
TRTGroupY1Y2Y3Y4Y5 * HLMstatusY5	31.66	1	31.66	4.675	.03	.00	4.67	.58
Error	11969.64	1767	6.77					
Total	78402.14	1771						
Corrected Total	12026.57	1770						

a. R Squared = .005 (Adjusted R Squared = .003)
b. Computed using alpha = .05

The analyses in tables 7.5 and 7.6 focus on the cross-sectional HLM with the ReadCAT Year1 variable as outcome. The results showed that those who were in this analysis were not significantly different (based on effect sizes) than those excluded from the HLM analysis and this difference did not depend on whether the youth was randomly assigned to the Read 180 or traditional English classroom. Further, there was no significant difference between the Read 180 and the Traditional groups with respect to the performance on the SRI at baseline.

Table 7.5. Descriptive Statistics by Treatment Groups and ReadCAT_Year 1 Cross-Sectional HLM Analysis Status: SRI baseline as Outcome

	ReadCAT HLM Model	Mean	SD	N
Read180	Out of the Analysis	769.27	199.15	700
	In the Analysis	731.47	190.49	133
	Total	763.24	198.16	833
Traditional	Out of the Analysis	788.73	188.56	661
	In the Analysis	750.16	200.64	110
	Total	783.23	190.68	771
Total	Out of the Analysis	778.72	194.25	1361
	In the Analysis	739.93	194.97	243
	Total	772.85	194.80	1604

Table 7.6. ReadCAT_Year1 Cross-Sectional Analysis of Variance Source Table: SRI baseline as Outcome

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	460015.17[a]	3	153338.39	4.06	.00	.00	12.19	.84
Intercept	4.726E8	1	4.726E8	12525.3	.00	.88	12525.37	1.00
TRTGroupY1Y2Y3Y4Y5	74450.32	1	74450.32	1.97	.16	.00	1.97	.28
XSec_Read_1Y	298274.13	1	298274.13	7.90	.00	.00	7.90	.82
TRTGroupY1Y2Y3Y4Y5 * XSec_Read_1Y	30.50	1	30.50	.00	.97	.00	.00	.05
Error	60367237.78	1600	37729.52					
Total	1.019E9	1604						
Corrected Total	60827252.96	1603						

a. R Squared = .008 (Adjusted R Squared = .006)
b. Computed using alpha = .05

Tables 7.7 and 7.8 present the comparable analysis using the Read CAT baseline as the dependent variable. The results showed that those who were in the analysis were not significantly different (based on effect sizes) than those excluded from the HLM analysis and this difference did not depend on whether the youth was randomly assigned to the Read 180 or traditional English classroom. Further, there was no significant difference between the Read 180 and the Traditional groups with respect to the performance on the ReadCAT at baseline.

Table 7.7. Descriptive Statistics by Treatment Groups and ReadCAT_Year 1 Cross-Sectional HLM Analysis Status: ReadCAT baseline as Outcome

	ReadCAT HLM Model	Mean	SD	N
Read180	Out of the Analysis	6.08	2.55	663
	In the Analysis	5.47	2.40	133
	Total	5.98	2.53	796
Traditional	Out of the Analysis	6.21	2.63	628
	In the Analysis	5.73	2.43	110
	Total	6.14	2.60	738
Total	Out of the Analysis	6.14	2.59	1291
	In the Analysis	5.59	2.41	243
	Total	6.05	2.57	1534

Table 7.8. ReadCAT_Year1 Cross-Sectional Analysis of Variance Source Table: ReadCAT baseline as Outcome

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	71.68[a]	3	23.89	3.64	.01	.00	10.92	.80
Intercept	27987.64	1	27987.64	4265.01	.00	.73	4265.01	1.00
TRTGroupY1Y2Y3Y4Y5	7.39	1	7.39	1.12	.28	.00	1.12	.18
XSec_Read_1Y	59.75	1	59.75	9.10	.00	.00	9.10	.85
TRTGroupY1Y2Y3Y4Y5 * XSec_Read_1Y	.79	1	.79	.12	.72	.00	.12	.06
Error	10040.08	1530	6.56					
Total	66330.20	1534						
Corrected Total	10111.76	1533						

a. R Squared = .007 (Adjusted R Squared = .005)

b. Computed using alpha = .05

The analyses in tables 7.9 and 7.10 focus on the Longitudinal HLM with the SRI variable as outcome. The results showed that those who were in this analysis were not significantly different (based on effect sizes) than those excluded from the HLM analysis and this difference did not depend on whether the youth was randomly assigned to the Read 180 or traditional English classroom. Further, there was no significant difference between the Read 180 and the Traditional groups with respect to the performance on the SRI at baseline.

Table 7.9. Descriptive Statistics by Treatment Groups and SRI Longitudinal HLM Analysis Status: SRI baseline as Outcome

	SRI Longitudinal HLM Analysis	Mean	SD	N
Read180	Out of the Analysis	730.28	229.86	92
	In the Analysis	767.33	193.65	741
	Total	763.24	198.16	833
Traditional	Out of the Analysis	779.97	195.66	119
	In the Analysis	783.82	189.90	652
	Total	783.23	190.68	771
Total	Out of the Analysis	758.31	212.17	211
	In the Analysis	775.05	192.01	1393
	Total	772.85	194.80	1604

Table 7.10. SRI Longitudinal Analysis of Variance Source Table: SRI baseline as Outcome

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	273863.70[a]	3	91287.90	2.41	.06	.00	7.23	.60
Intercept	4.230E8	1	4.230E8	11177.0	.00	.87	11177.08	1.00
TRTGroupY1Y2Y3Y4Y5	197733.04	1	197733.04	5.22	.02	.00	5.22	.62
Long_ITT_Flag	75474.21	1	75474.21	1.99	.15	.00	1.99	.29
TRTGroupY1Y2Y3Y4Y5 * Long_ITT_Flag	49734.06	1	49734.06	1.31	.25	.00	1.31	.20
Error	60553389.25	1600	37845.86					
Total	1.019E9	1604						
Corrected Total	60827252.	1603						

a. R Squared = .005 (Adjusted R Squared = .003)

b. Computed using alpha = .05

Tables 7.11 and 7.12 present the comparable analysis using the ReadCAT baseline as the dependent variable. The results showed that those who were in the analysis were not significantly different (based on effect sizes) than those excluded from the HLM analysis and this difference did not depend on whether the youth was randomly assigned to the Read 180 or traditional English classroom. Further, there was no significant difference between the Read 180 and the Traditional groups with respect to the performance on the ReadCAT at baseline.

Table 7.11. Descriptive Statistics by Treatment Groups and SRI Longitudinal HLM Analysis Status: ReadCAT baseline as Outcome

	SRI Longitudinal HLM Analysis	Mean	SD	N
Read180	Out of the Analysis	5.97	2.51	55
	In the Analysis	5.98	2.54	741
	Total	5.98	2.53	796
Traditional	Out of the Analysis	6.13	2.59	86
	In the Analysis	6.14	2.61	652
	Total	6.14	2.60	738
Total	Out of the Analysis	6.07	2.55	141
	In the Analysis	6.05	2.57	1393
	Total	6	3	1534

Table 7.12. SRI Longitudinal Analysis of Variance Source Table: ReadCAT baseline as Outcome

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[b]
Corrected Model	9.53[a]	3	3.17	.48	.69	.00	1.44	.14
Intercept	17937.61	1	17937.61	2716.68	.00	.64	2716.68	1.00
TRTGroupY1Y2Y3Y4Y5	2.99	1	2.99	.45	.50	.00	.45	.10
Long_ITT_Flag	.00	1	.00	.00	.97	.00	.00	.05
TRTGroupY1Y2Y3Y4Y5 * Long_ITT_Flag	.00	1	.00	.00	.99	.00	.00	.05
Error	10102.23	1530	6.60					
Total	66330.20	1534						
Corrected Total	10111.76	1533						

a. R Squared = .001 (Adjusted R Squared = -.001)

b. Computed using alpha = .05

Because of the issues associated with the dataset used to model the ReadCAT (last score), we chose not to run a test of equivalence on that data set.