



**IMPROVING STUDENT LITERACY
IN THE PHOENIX UNION HIGH SCHOOL DISTRICT
2003-04 AND 2004-05
Final Report**

RICHARD N. WHITE
M. BRUCE HASLAM
GINA M. HEWES

July 7, 2006

Policy Studies Associates, Inc.
1718 Connecticut Avenue, NW
Suite 400
Washington, DC 20009

This report was prepared under a contract from Scholastic, Inc. The findings and conclusions are those of the authors. They do not necessarily represent the opinions of Scholastic or the Phoenix Union High School District, and no endorsement should be inferred.

Contents

	Page
Summary of Findings	i
Introduction	1
Cohort 1: Students in Ninth Grade in 2003-04	5
Cohort 2: Students in Ninth Grade in 2004-05	19
Cohort 3: Students in Tenth Grade in 2003-04	25
Cohort 4: Students in Tenth Grade in 2004-05	27
Conclusions and Suggestions for Additional Research	33
Appendix A: Using Normal Curve Equivalent Scores	A-1
Appendix B: Propensity Matching Procedures for Identifying Matched Comparison Students	B-1
Appendix C: Central Tendencies and Distributions	C-1
Appendix D: Crosswalk between SRI Lexile Scale Scores and Normal Curve Equivalent Scores	D-1

List of Exhibits

Page

Summary of Findings

Exhibit A: Data Available by Cohort	ii
Exhibit B: Cohort 1 2003 and 2004 SAT9 Reading Comprehension Subtest Scores.....	iv
Exhibit C: Cohort 2 2005 TerraNova Reading Scores.....	v
Exhibit D: Cohort 1 ELL Students' AIMS Reading Performance Levels	vi
Exhibit E: Cohort 1 Distribution of Changes in SRI Scores (in NCEs)	vii
Exhibit F: Cohort 2 Distribution of Changes in SRI Scores (in NCEs)	viii

Introduction

Exhibit 1: Data Available by Cohort.....	2
--	---

Cohort 1: Students in Ninth Grade in 2003-04

Exhibit 2: Cohort 1 Average SAT9 Reading Comprehension Subtest Scores (in NCEs): 2003 and 2004	6
Exhibit 3: Average Scale Scores on the Tenth-Grade AIMS Reading Test: Cohort 1 Participants and Nonparticipants	8
Exhibit 4: Proportion of Cohort 1 Participants and Matched Nonparticipants Who Met or Exceeded the State Proficiency Standard on the Tenth-Grade AIMS Reading Test.....	9
Exhibit 5: Cohort 1 SRI Scores at the Initial and Final Test Points (in NCEs). 10	
Exhibit 6: Cohort 1 Changes in SRI Scores (in NCEs)	10
Exhibit 7: Cohort 1 Performance on the Tenth-Grade AIMS Reading Test, by Gains on SRI.....	12
Exhibit 8: Changes in SAT9 Reading Comprehension Subtest Scores (in NCEs): Cohort 1 ELL Participants and Matched Nonparticipants.....	13

List of Exhibits (continued)

	Page
Exhibit 9: Cohort 1 AIMS Reading Performance Levels: ELL Participants and Matched Nonparticipants	14
Exhibit 10: Cohort 1 Changes in SAT9 Reading Comprehension Subtest Scores (in NCEs): Participants and Matched Nonparticipants	16
Cohort 2: Students in Ninth Grade in 2004-05	
Exhibit 11: Cohort 2 Scores on Eighth-Grade SAT9 Reading Comprehension Subtest and Ninth-Grade TerraNova Reading Test (in NCEs)	19
Exhibit 12: Cohort 2 SRI Scores at the Initial and Final Test Points (in NCEs) .	20
Exhibit 13: Cohort 2 Changes in SRI Scores.....	21
Cohort 3: Students in Tenth Grade in 2003-04	
Exhibit 14: Cohort 3 SRI Scores at the Initial and Final Test Points (in NCEs) .	23
Exhibit 15: Cohort 3 Changes in SRI Scores.....	24
Cohort 4: Students in Tenth Grade in 2004-05	
Exhibit 16: Cohort 4 SRI Scores at the Initial and Final Test Points (in NCEs) .	26
Exhibit 17: Cohort 4 Changes in SRI Scores.....	27
Exhibit 18: Cohort 4 Initial SRI Scores (in NCEs): New and Continuing Participants	28
Exhibit 19: Cohort 4 Changes in SRI Scores (in NCEs): New and Continuing Participants	29
Exhibit 20: Cohort 4 SRI Scores by Date and Grade of Participation.....	30
Appendix A: Using Normal Curve Equivalent Scores	
Exhibit A-1: Distribution of Student Scores, Normal Curve Equivalent Scores, and Percentiles.....	A-1

List of Exhibits (continued)

Page

Appendix B: Propensity Matching Procedures for Identifying Matched Comparison Students

Exhibit B-1: Characteristics of READ 180 Participants and Matched Nonparticipants..... B-2

Appendix C: Central Tendencies and Distributions

Exhibit C-1: Cohort 1 Reading Achievement Outcomes (in NCEs), by Measure and Subgroup C-1

Exhibit C-2: Cohort 2 Reading Achievement Outcomes (in NCEs), by Measure and Subgroup C-3

Exhibit C-3: Cohort 3 Reading Achievement Outcomes (in NCEs)..... C-4

Exhibit C-4: Cohort 4 Reading Achievement Outcomes (in NCEs)..... C-4

Appendix D: Crosswalk between SRI Lexile Scale Scores and Normal Curve Equivalent Scores

Exhibit D-1: Crosswalk between SRI Lexile Scale Scores and Normal Curve Equivalent Scores D-1

Summary of Findings

In fall 2003, the Phoenix (Arizona) Union High School District (PUHSD) began using Scholastic's READ 180 program, Stage C, Version 1.6, to help struggling ninth- and tenth-graders become proficient readers and improve their opportunities for success in school. To gauge the impact of READ 180 among PUHSD ninth-graders and tenth-graders during the first two years of implementation, Scholastic asked Policy Studies Associates (PSA) to analyze data from a variety of reading assessments administered as part of the state assessment system or as part of the implementation of READ 180. Data were available for four cohorts of PUHSD students, with the cohorts defined in terms of when students participated in READ 180.

- Cohort 1 includes students who participated in READ 180 as ninth-graders in 2003-04 and matched nonparticipants
- Cohort 2 includes students who participated in READ 180 as ninth-graders in 2004-05 and matched nonparticipants
- Cohort 3 includes tenth-graders who participated in READ 180 for one semester in 2003-04
- Cohort 4 includes tenth-graders who participated in READ 180 for one semester in 2004-05

It should be noted that Cohorts 1 and 4 overlap to some degree, with the result that Cohort 4 includes two subgroups of students: students who were continuing in READ 180 from ninth grade and students whose first participation in READ 180 was in the tenth grade.

PSA's analysis addresses three questions:

- How did the changes in reading proficiency achieved by READ 180 students compare with the changes achieved by matched nonparticipants?
- Did gains in reading proficiency achieved by ninth-grade READ 180 participants persist into the tenth grade?
- What gains did students in each cohort make on the Scholastic Reading Inventory (SRI) and what does the magnitude of gains on the SRI indicate about potential gains on state assessments?

The data available for this study are displayed in Exhibit A.

Exhibit A
Data Available by Cohort

Data	Cohort 1 (Ninth-graders in 2003-04)		Cohort 2 (Ninth-graders in 2004-05)		Cohort 3 (Tenth-graders in 2003-04)		Cohort 4 (Tenth-graders in 2004-05)	
	READ 180 Participants	Nonparticipants	READ 180 Participants	Nonparticipants	READ 180 Participants	Nonparticipants	READ 180 Participants	Nonparticipants
Dates of Enrollment in READ 180	X		X		X		X	
SRI Scores	X		X		X		X	
Eighth-Grade SAT9 Reading Comprehension Scores	X	X	X	X				
Ninth-Grade SAT9 Reading Comprehension Scores	X	X						
Ninth-Grade TerraNova Reading Scores			X	X				
Tenth-Grade AIMS Reading Score*	X	X						
ELL Status	X	X	X	X				
Special Education Status	X	X	X	X				

*AIMS is the Arizona Instrument to Measure Standards.

These data permit several kinds of analysis of the impact of READ 180 on student reading proficiency, including:

- Analysis that compares Cohort 1 scores on the SAT9 Reading Comprehension Subtest in eighth grade and ninth grade with scores for a matched sample of nonparticipants
- Analysis that compares outcomes on the AIMS Reading test achieved by Cohort 1 READ 180 participants in the tenth grade with outcomes for matched nonparticipants

- Analysis of participants' initial reading proficiency and gains in proficiency while participating in READ 180, as measured by the SRI
- Analysis that examines changes in reading proficiency, among all four cohorts of participants, as measured by the SRI, and exploration of the relationship between gains in SRI scores and gains in scores on the end-of-year state assessments in reading
- Analysis that examines changes in reading proficiency achieved by subgroups, such as students identified as English Language Learners (ELL) or by their performance on baseline assessments

Three Key Findings

Overall the analysis of the PUHSD data suggests three key findings about the impact of participation in READ 180 on students' reading proficiency:

1. ***READ 180 students in Cohorts 1 and 2 outperformed matched comparison groups of nonparticipants on tests included as part of the Arizona state assessment system.*** More specific findings include the following:
 - Cohort 1 participants and a matched comparison group of nonparticipants scored lower on the 2004 SAT9 Reading Comprehension Subtest than they did in 2003. However, Cohort 1 READ 180 students experienced smaller decreases (a decrease of 0.6 NCEs versus a decrease of 1.9 NCEs, with the difference being statistically significant) (Exhibit B).
 - Overall, all 2003-04 PUHSD ninth-graders achieved lower scores on their ninth-grade SAT9 Reading Comprehension Subtest than they had on their eighth-grade test (a decline from 41.8 NCEs to 36.0 NCEs, with the decline representing a statistically significant difference). (This decline is depicted by the red line in Exhibit B.)

Exhibit B Cohort 1 2003 and 2004 SAT9 Reading Comprehension Subtest Scores

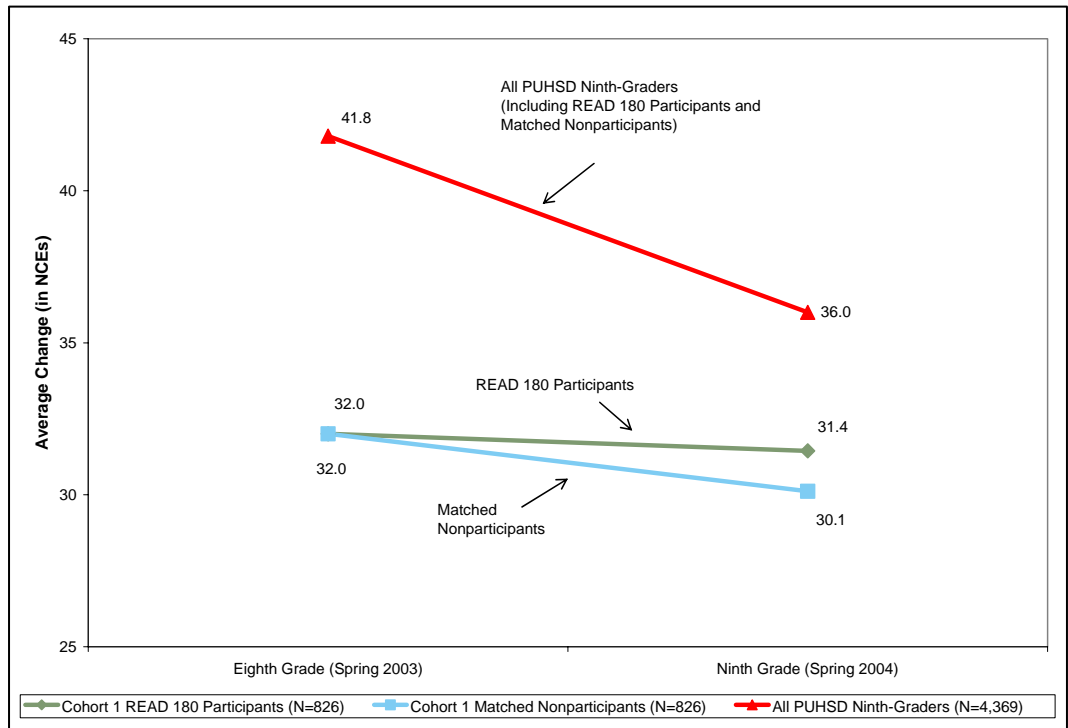


Exhibit reads: Cohort 1 participants averaged 32.0 NCEs (± 0.3 NCE) on the spring 2003 SAT9 Reading Comprehension Subtest and 31.4 NCEs (± 0.3 NCE) on the spring 2004 test.

- Cohort 2 participants achieved higher average scores on the TerraNova Reading test than did a matched comparison group of nonparticipants (41 NCEs versus 38 NCEs, with the difference being statistically significant) (Exhibit C).

Exhibit C Cohort 2 2005 TerraNova Reading Scores

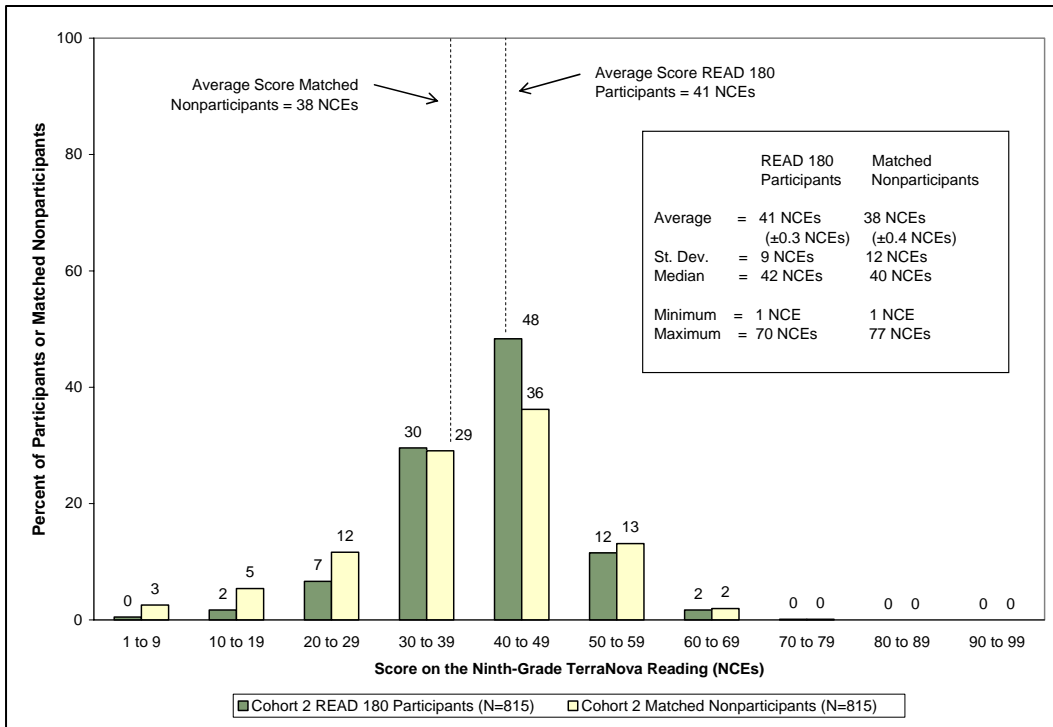


Exhibit reads: Less than 1 percent of Cohort 2 participants scored between 1 and 9 NCEs on the spring 2005 TerraNova Reading test.

- Cohort 2 participants who were eligible for ELL services achieved higher scores on the 2005 TerraNova Reading test than did matched nonparticipants (39.9 NCEs versus 35.4 NCEs, with the difference being statistically significant).
 - Cohort 2 participants who had scored 40 NCEs or below on their eighth-grade SAT9 Reading Comprehension Subtest achieved higher scores on the 2005 TerraNova Reading Test than did matched nonparticipants (39.8 NCEs versus 36.2 NCEs, with the difference being statistically significant).
2. *Overall, Cohort 1 participants performed about the same as nonparticipants on the tenth-grade AIMS Reading test as tenth-graders (average 664.1 scale-score points versus 664.2 scale-score points). However, results on the AIMS Reading test indicate that the benefits in reading proficiency associated with participation in READ 180 persist into the next school year for several categories of students.*

- Cohort 1 ELL participants outperformed matched nonparticipants on the AIMS Reading test (654 scale-score points versus 646 scale-score points, with the difference being statistically significant). Correspondingly, a larger proportion of READ 180 ELL participants than matched nonparticipants scored at the “meets or exceeds standard” level in tenth-grade reading, and a larger proportion scored at the “approaching standard” level (Exhibit D).

Exhibit D
Cohort 1 ELL Students’ AIMS Reading Performance Levels

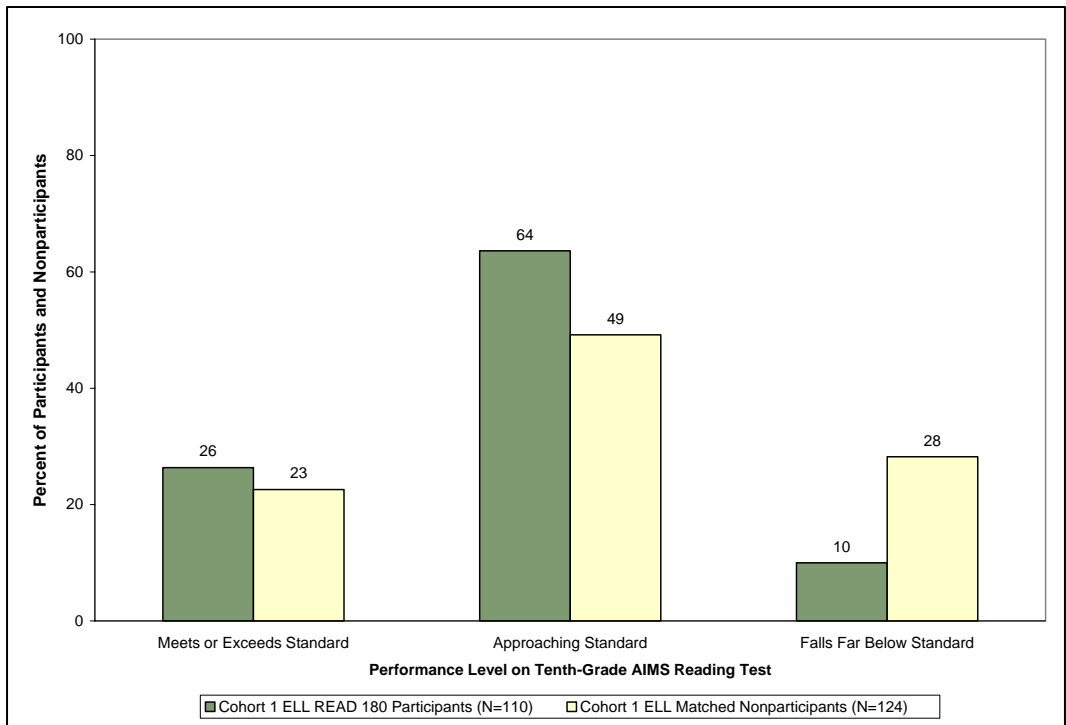


Exhibit reads: Twenty-six percent of Cohort 1 ELL participants scored at the “Meets or exceeds the standard” performance level on the tenth-grade AIMS Reading test.

- Cohort 1 participants who scored below 35 NCEs on their eighth-grade SAT9 Reading Comprehension Subtest outperformed matched nonparticipants on the AIMS Reading test (656 scale-score points versus 651 scale-score points, with the difference being statistically significant).

- Cohort 1 participants who had achieved gains of 10 NCEs or more on the SRI attained higher scores and proficiency levels on the tenth-grade AIMS Reading test than did other former participants. Cohort 1 participants who achieved gains of 10 NCEs or more on the SRI averaged 667 scale-score points on the tenth-grade AIMS Reading test compared with an average of 661 scale-score points among Cohort 1 participants who had gained less than 10 NCEs (with this difference being statistically significant).

3. ***READ 180 participants made substantial gains on the SRI assessment of reading proficiency.***

- Participants in Cohorts 1 and 2 averaged gains of 14 NCEs and 11 NCEs, respectively (Exhibits E and F).
- Participants in Cohorts 3 and 4 averaged gains of 7.5 NCEs and 1.5 NCEs respectively, after one semester of participation in the program.

Exhibit E
Cohort 1 Distribution of Changes in SRI Scores (in NCEs)

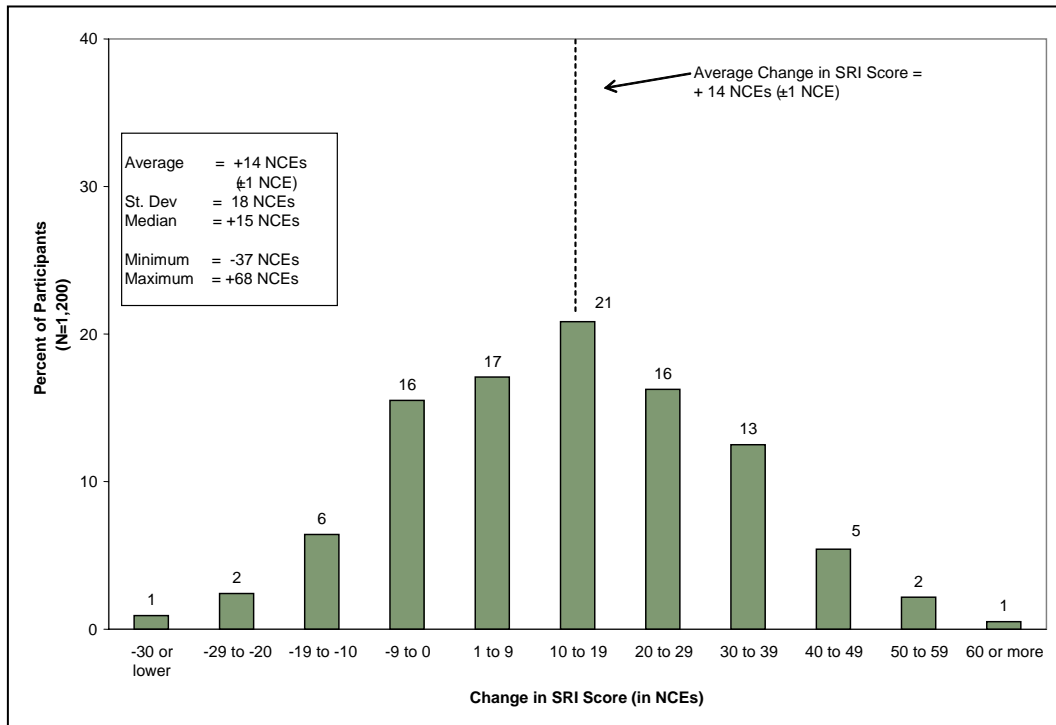


Exhibit reads: The average change in SRI scores between the initial and final test points is a gain of 14 NCEs.

Exhibit F Cohort 2 Distribution of Changes in SRI Scores (in NCEs)

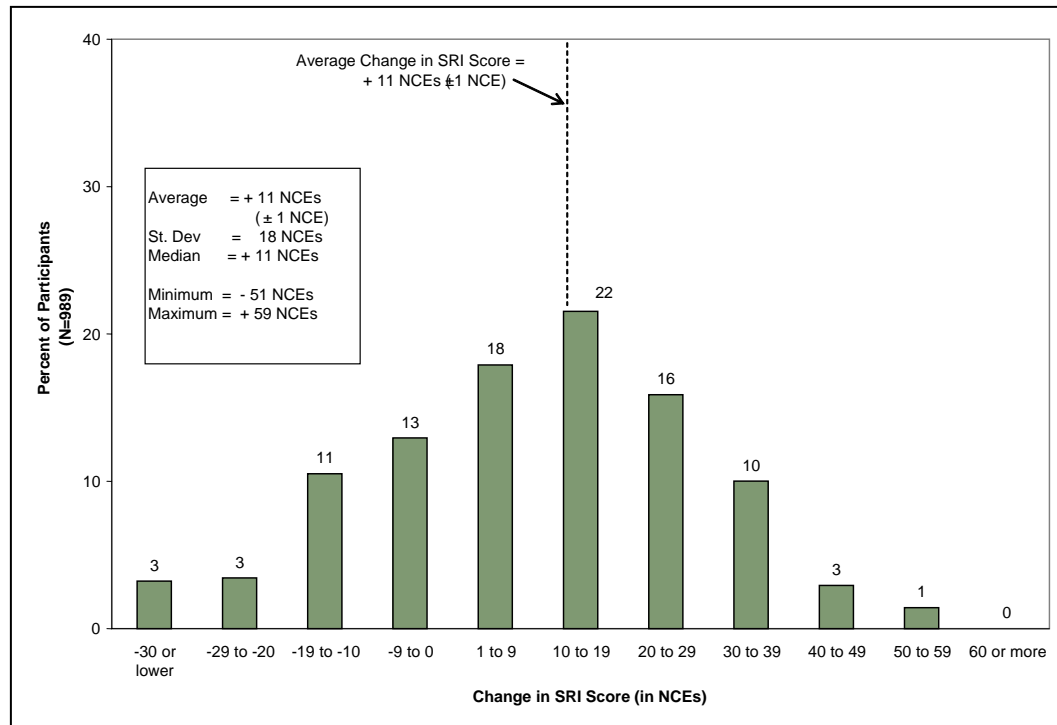


Exhibit reads: The average gain in SRI scores between the initial and final test points was a gain of 11 NCEs.

Conclusions and Suggestions for Additional Research

This study of four cohorts of PUHSD students in 2003-04 and 2004-05 found that participation in READ 180 was associated with meaningful benefits in reading proficiency.

The performance of 2003-04 READ 180 participants on the tests administered as part of the Arizona state assessment showed benefits in terms of gains in reading proficiency. The SAT9 Reading Comprehension Subtest scores of Cohort 1 participants decreased less between eighth and ninth grades than did the scores of a group of similar nonparticipants who were matched on the bases of eighth-grade test scores and demographic characteristics. The differences between the participants and matched nonparticipants are statistically significant. However, there are no overall differences between Cohort 1 participants and matched nonparticipants on the tenth-grade AIMS Reading Test.

Analysis of trends in the reading proficiency of Cohort 2 participants was complicated by the change in the standardized test that was administered. These students completed the SAT9 Reading Comprehension Subtest in eighth grade and the TerraNova Reading test in ninth grade. READ 180 participants averaged

higher scores on the ninth-grade TerraNova Reading Test than did the matched group of nonparticipants.

Among Cohort 1 and Cohort 2 students, two subgroups of students benefited more from participation in READ 180. READ 180 participants who were eligible for ELL services achieved larger gains on the state assessments than did the matched nonparticipants who were eligible for ELL. These differences were observed in changes on the state assessments between eighth grade and ninth grade, and again in the differences in the average scores achieved by participants and matched nonparticipants on the tenth-grade reading tests.

The second group of students who benefited more from READ 180 were students who had lower levels of reading proficiency prior to participating in the program. For example, 2004-05 ninth-grade READ 180 participants who had scored 40 NCEs or below on their eighth-grade SAT9 Reading Comprehension Subtest scored higher on their ninth-grade TerraNova Reading Test than did the matched group of nonparticipants.

Tenth-grade students who participated in READ 180 for one semester in 2003-04 averaged SRI gains of more than 5 NCEs. Gains among the 2004-05 cohort of tenth-grade participants were smaller.

The findings from this study also suggest several areas for additional research and analysis to improve understanding of the relationship between participation in READ 180 and academic performance.

- ***A broader range of academic outcomes associated with participation in READ 180.*** In addition to changes in reading proficiency as measured by the SRI, TerraNova Reading Test, and AIMS Reading Test, it would be useful to explore whether there is any relationship with indicators of attachment to school and progress toward high school graduation, including school attendance, suspensions/expulsions, course grades, course credits earned, and grade promotion/retention.
- ***The relationship between the quality of implementation and sustained use of READ 180 and changes in reading proficiency.*** Examining the extent to which varying amounts of gains in student reading proficiency are associated with key features of program implementation, including implementation of the program's complex instructional model, would help identify needs for enhanced implementation and teacher professional development.
- ***The collection and analysis of data for additional years and cohorts of students.*** The comparison of the performance of the Cohort 1 READ 180 participants and matched participants on their

eleventh-grade AIMS tests and other areas of academic performance would yield additional information on the persistence of the relationship between participation in READ 180 and academic performance. Analysis of the performance of the Cohort 2 READ 180 participants and matched nonparticipants on their tenth-grade AIMS Reading Tests and other areas of academic performance would strengthen the findings about the Cohort 1 students presented in this report. Analysis of the results for the ninth-grade students enrolled in READ 180 during 2005-06 would allow the measure of change in reading proficiency between eighth and ninth grades on the TerraNova Reading Test. This would allow the exploration of whether the relationship between participation and gains on the SAT9 Reading Comprehension Subtest were also to be found on the TerraNova Reading Test.

Introduction

In fall 2003, the Phoenix (Arizona) Union High School District (PUHSD) began using Scholastic's READ 180 program, Stage C, Version 1.6, to help struggling ninth- and tenth-graders become proficient readers and improve their opportunities for success in school. The district targeted students who were reading one or more grade levels below their assigned grade levels to participate in the program. Thus, all ninth-graders who were reading below a grade equivalent of 8.0 (i.e., at a level equivalent to the average student nationwide entering eighth grade), as measured on the SAT9 administered in the spring of their eighth-grade year, were eligible to participate for a full school year. Similarly, tenth-graders who were reading below a grade equivalent of 9.0 on the SAT9 in the spring of their ninth-grade year were eligible to participate in the program, but for only one semester, either the fall or the spring.

To gauge the impact of READ 180 among PUHSD ninth-graders and tenth-graders during the first two years of implementation, Scholastic asked Policy Studies Associates (PSA) to analyze data from a variety of reading assessments administered as part of the state assessment system or as part of the implementation of READ 180. Data were available for four cohorts of PUHSD students, with the cohorts defined in terms of when students participated in READ 180.

- Cohort 1 includes students who participated in READ 180 as ninth-graders in 2003-04 and matched nonparticipants
- Cohort 2 includes students who participated in READ 180 as ninth-graders in 2004-05 and matched nonparticipants
- Cohort 3 includes tenth-graders who participated in READ 180 for one semester in 2003-04
- Cohort 4 includes tenth-graders who participated in READ 180 for one semester in 2004-05

It should be noted that Cohorts 1 and 4 overlap to some degree, with the result that Cohort 4 includes two subgroups of students: students who were continuing in READ 180 from ninth grade and students whose first participation in READ 180 was in the tenth grade.

PSA's analysis addresses three questions:

- How did the changes in reading proficiency achieved by READ 180 students compare with the changes achieved by matched nonparticipants?

- Did gains in reading proficiency achieved by ninth-grade READ 180 participants persist into the tenth grade?
- What gains did students in each cohort make on the Scholastic Reading Inventory (SRI)¹ and what does the magnitude of gains on the SRI indicate about potential gains on state assessments?

The data available for this study are displayed in Exhibit 1.

**Exhibit 1
Data Available by Cohort**

Data	Cohort 1 (Ninth-graders in 2003-04)		Cohort 2 (Ninth-graders in 2004-05)		Cohort 3 (Tenth-graders in 2003-04)		Cohort 4 (Tenth-graders in 2004-05)	
	READ 180 Participants	Nonparticipants	READ 180 Participants	Nonparticipants	READ 180 Participants	Nonparticipants	READ 180 Participants	Nonparticipants
Dates of Enrollment in READ 180	X		X		X		X	
SRI Scores	X		X		X		X	
Eighth-Grade SAT9 Reading Comprehension Scores	X	X	X	X				
Ninth-Grade SAT9 Reading Comprehension Scores	X	X						
Ninth-Grade TerraNova Reading Scores			X	X				
Tenth-Grade AIMS Reading Score*	X	X						
ELL Status	X	X	X	X				
Special Education Status	X	X	X	X				

*AIMS is the Arizona Instrument to Measure Standards.

¹ The Scholastic Reading Inventory (SRI) is a reading comprehension test administered to students participating in READ 180 up to five times per year. The results are used by READ 180 teachers to assess students' reading levels, track students' reading growth over time, match students to appropriate reading materials, and help plan instruction according to students' needs.

These data permit several kinds of analysis of the impact of READ 180 on student reading proficiency, including:

- Analysis that compares Cohort 1 scores on the SAT9 Reading Comprehension Subtest in eighth grade and ninth grade with scores for a matched sample of nonparticipants
- Analysis that compares outcomes on the AIMS Reading test achieved by Cohort 1 READ 180 participants in the tenth grade with outcomes for matched nonparticipants
- Analysis of participants' initial reading proficiency and gains in proficiency while participating in READ 180, as measured by the SRI
- Analysis that examines changes in reading proficiency, among all four cohorts of participants, as measured by the SRI, and exploration of the relationship between gains in SRI scores and gains in scores on the end-of-year state assessments in reading
- Analysis that examines changes in reading proficiency achieved by subgroups, such as students identified as English Language Learners (ELL) or by their performance on baseline assessments

Data that were available also preclude certain kinds of analyses:

- Analysis that examine Cohort 2 participants' *changes* in reading proficiency between eighth grade and ninth grade, because the standardized test administered during spring 2005 changed from the SAT9 Reading Comprehension Subtest to the TerraNova Reading test
- Analysis that compare Cohort 2 participants' and nonparticipants' tenth-grade reading scores, because the spring 2006 tenth-grade AIMS Reading test scores were not available
- Analysis that examines Cohort 3 and Cohort 4 participants' and nonparticipants' gains in reading proficiency, as measured by the AIMS Reading test, because these scores were not available

Before discussing the results of the analysis conducted for this study, a note about the comparison groups is in order. Because the major criterion for selecting students to participate in READ 180 was below grade-level performance on their eighth-grade or ninth-grade SAT9 Reading Comprehension Subtest, comparing changes in participants' reading proficiency with changes among all nonparticipants could be misleading. For example, the students selected to

participate in READ 180 averaged a score of 32.0 normal curve equivalents (NCEs)² on the eighth-grade SAT9 Reading Comprehension Subtest and only one percent scored above 50 NCEs, a commonly used marker of performance at grade level. Among all other PUHSD eighth-graders, the average score was 44.0 NCEs, and 40 percent scored above 50 NCEs. To avoid the problem of incomparable starting points, a propensity matching procedure was used to identify a subset of nonparticipants whose reading level and ELL eligibility were similar to those of participants.³

The next four sections of the report examine, in turn, the changes in reading proficiency achieved by each of the four cohorts of students. Each section focuses on the results of the analysis of the outcome data that were available for the cohort and, as appropriate, examines the results for student subgroups. The section on Cohort 1 is the longest and fullest because this is the cohort for which the most data were available. Sections on Cohorts 3 and 4 are short because these are the cohorts for which the least data were available. The last section presents concluding observations and recommendations for additional research. Four appendices provide additional technical information, including (1) an explanation of the use of NCEs to report on changes in student reading proficiency; (2) a description of propensity matching, how it works, and what the matching produced for this study in particular; (3) a full set of means, standard deviations, and other measures of central tendency and distribution for the achievement measures used in this study; and (4) a crosswalk between Lexile scores and NCEs.

² See Appendix A for a discussion of the use of NCEs to analyze and report on changes in reading proficiency.

³ See Appendix B for a description of propensity matching.

Cohort 1: Students in Ninth Grade in 2003-04

This section examines the changes in reading proficiency achieved by PUHSD ninth-graders, both READ 180 participants who participated in READ 180 for a full year in 2003-04, the first year of implementation in the district, and matched nonparticipants. In preview, the analysis of changes in reading proficiency achieved by Cohort 1 students found that:

- *READ 180 participants outperformed matched nonparticipants on the ninth-grade SAT9 Reading Comprehension Subtest*
- *ELL participants outperformed matched nonparticipants*
- *Participants with lower scores at the eighth-grade test point tended to achieve greater gains on the ninth grade than participants with higher initial scores, and greater gains than matched nonparticipants*

This section begins with a comparison of scores achieved by READ 180 participants and those achieved by a matched group of nonparticipants on the SAT9 Reading Comprehension Subtest. After reviewing the SAT9 Reading Comprehension Subtest results of Cohort 1 students as ninth-graders, this section focuses on their performance on the AIMS Reading Test as tenth-graders. Next, it investigates SRI outcomes for Cohort 1 students and how SRI scores relate to the other achievement measures. Finally, this section explores the performance of two subgroups: ELL students and students scoring at different levels of reading proficiency prior to participation in READ 180.

Cohort 1 Students' Performance on SAT9 Reading Comprehension Subtest

Prior to the 2004-05 school year, all PUHSD students completed the SAT9 Reading Comprehension Subtest at the end of eighth grade and again at the end of ninth grade. Consequently, scores from the SAT9 Reading Comprehension Subtest administered in spring 2003 when Cohort 1 students were eighth-graders, and spring 2004 when the same students were in the ninth grade, were available for analysis of the changes in reading proficiency achieved by Cohort 1 students.

The average score on the SAT9 Reading Comprehension Subtest administered to all PUHSD ninth-graders in spring 2004 declined from the previous year (a decline in NCEs from 41.8 to 36.0, with the decline representing a statistically significant difference, as determined by a paired-samples t-test and $p < .05$). This overall decline is depicted by the red line in Exhibit 2.

Exhibit 2 Cohort 1 Average SAT9 Reading Comprehension Subtest Scores (in NCEs): 2003 and 2004

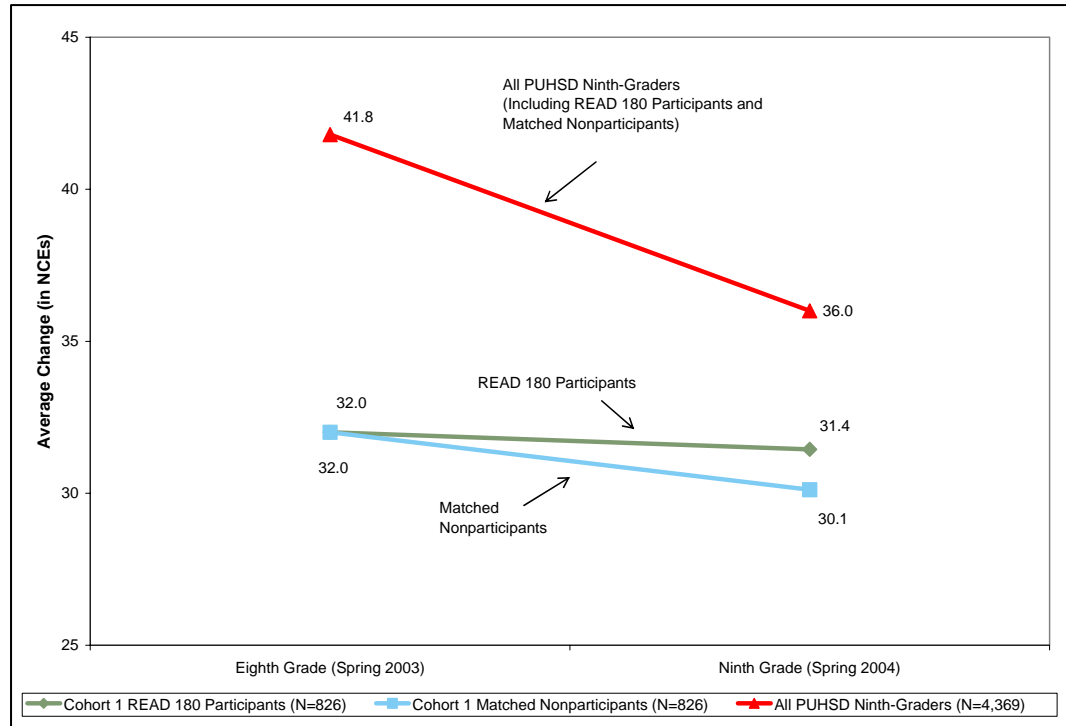


Exhibit reads: Cohort 1 READ 180 participants averaged scores of 32.0 NCEs (± 0.3 NCE) on the spring 2003 SAT9 Reading Comprehension Subtest and 31.4 NCEs (± 0.3 NCE) at the spring 2004 test point.⁴

Within the context of the overall decrease, a comparison of the SAT9 scores achieved by READ 180 students and their matched nonparticipants reveals that Cohort 1 participants averaged a *smaller decrease* in scores than did the matched nonparticipants. Specifically, Cohort 1 participants averaged a score of 32.0 NCEs (± 0.3 NCE) at the spring 2003 test point, with this average dipping to 31.4 NCEs (± 0.3 NCE) at the spring 2004 test point (a decrease of 0.6 NCEs). During the same period the average score of the matched nonparticipants dropped from 32.0 NCEs (± 0.3 NCE) to 30.1 NCEs (± 0.4 NCE) (a decrease of 1.9 NCEs).⁵ The difference between the two groups of students at the second test

⁴ Numbers with the \pm symbol indicate plus and minus margins of error, measured in this study as the 95 percent confidence interval around a mean or other point estimate.

⁵ See Appendix C for additional details on means, standard deviations, and other measures of central tendency and distribution.

point test is statistically significant and represents an effect size of +0.23.⁶ This effect size can also be understood to indicate that 59 percent of the READ 180 students scored above the comparison group's mean score on the SAT9 Reading Comprehension Subtest.

Cohort 1 Performance on the Tenth-Grade AIMS Reading Test

Like all PUHSD tenth-graders, Cohort 1 participants and matched nonparticipants completed the AIMS Reading Test in the spring of tenth grade (spring 2005). However, because this test measures different constructs than the SAT9 Reading Comprehension Subtest Test, and the scoring of the test is based on different norming populations, it is not possible to directly compute the change in reading proficiency between test points. Therefore, analysis of performance on this test is limited to comparing the scores achieved on the tenth-grade test by Cohort 1 participants and matched nonparticipants.

The results of the comparison of the tenth-grade AIMS Reading Test results indicates that average scores achieved by the READ 180 participants and the matched nonparticipants were almost identical- an average of 664.1 scale score points for the participants versus an average of 664.2 scale score points for the nonparticipants (Exhibit 3, right-hand panel).⁷ The difference in scores is not statistically significant.⁸

An ANCOVA analysis was used to estimate the effects of participation in READ 180, controlling for students' reading proficiency as measured by their score on the eighth-grade SAT9 Reading Comprehension Subtest. The adjusted means for Cohort 1 READ 180 participants and matched nonparticipants were 683.0 and 684.5 scale-score points, respectively (Exhibit 3, left-hand panel). The difference in the adjusted means is not statistically significant.

⁶ An effect size is useful for estimating the size or importance of the program effect, represented here as differences in the change in test score between the participants and nonparticipants. In this case, the effect size was estimated by subtracting the difference between the average individual change between participants' eighth-grade and ninth-grade SAT9 Reading Comprehension Subtest and the average change among the matched nonparticipants, and dividing the result by the pooled standard deviation in the gains. For this and other analyses of changes in reading proficiency, we have adopted the standard of an effect size of 0.1 representing a small effect size. (Note that "effect size" is a conventional research term, and is not necessarily meant to imply causality.)

⁷ Scores on the eighth-grade SAT9 Reading Comprehension Subtest administered in spring 2003 and on the AIMS Reading test administered in spring 2005 were available for 724 participants and 724 matched nonparticipants.

⁸ Unless otherwise noted, statistical significance was determined with the ANOVA procedure. A p value of less than .05 was used as the threshold of significance.

Exhibit 3
Average Scale Scores on the Tenth-Grade AIMS Reading Test:
Cohort 1 Participants and Nonparticipants

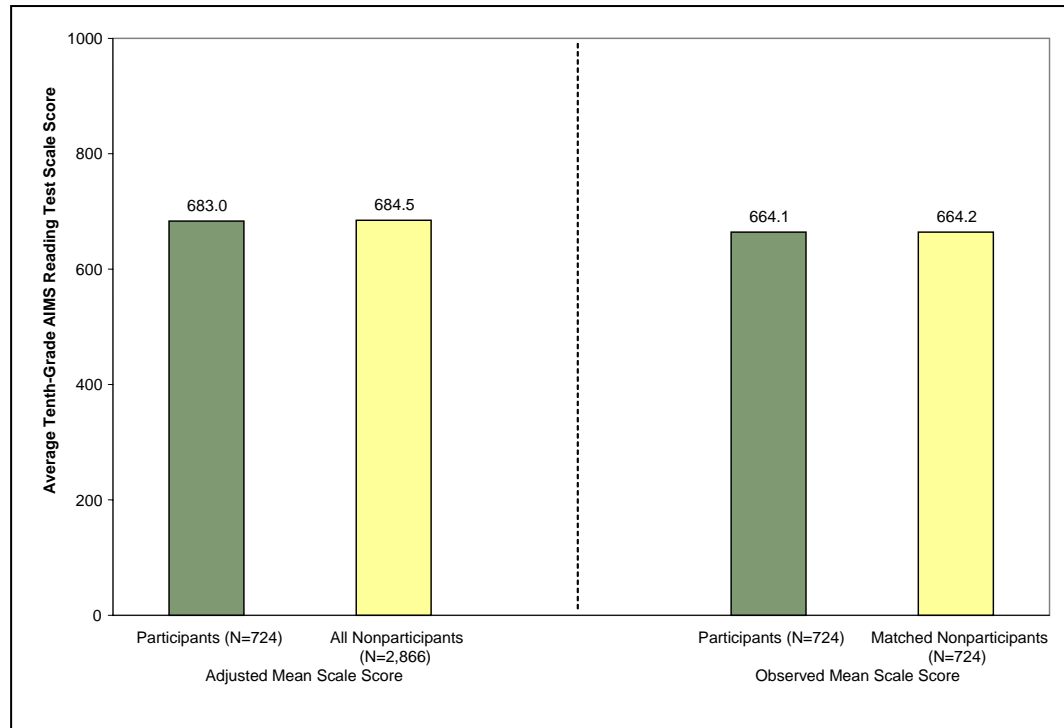


Exhibit reads: The adjusted mean AIMS Reading Test scale score for all Cohort 1 participants and 683.0 scale-score points, compared with an adjusted mean of 684.5 scale-score points for all nonparticipating students in Cohort 1. This indicates that after controlling for performance on the eighth-grade SAT9 Reading Comprehension Subtest, the students selected to participate in READ 180 were predicted to score slightly lower than all other PUHSD ninth-grade students, and the difference was not statistically significant.

Using the state accountability system as a lens to investigate the changes in reading proficiency achieved by Cohort 1 participants and nonparticipants, the analysis compared their respective rates of meeting or exceeding the state standard for reading proficiency. The results show that participants were as likely as matched nonparticipants to perform at the “meets or exceeds the standard” level on the AIMS Reading test (Exhibit 4). Specifically, 41 percent of Cohort 1 participants versus 43 percent of nonparticipants performed at this level, but the difference is not statistically significant ($p=.52$, chi-square test).

As will be discussed in more detail after reviewing SRI results for Cohort 1 participants, the Cohort 1 participant subgroups achieved higher scores than did matched nonparticipants.

Exhibit 4
Proportion of Cohort 1 Participants and Matched Nonparticipants
Who Met or Exceeded the State Proficiency Standard
on the Tenth-Grade AIMS Reading Test

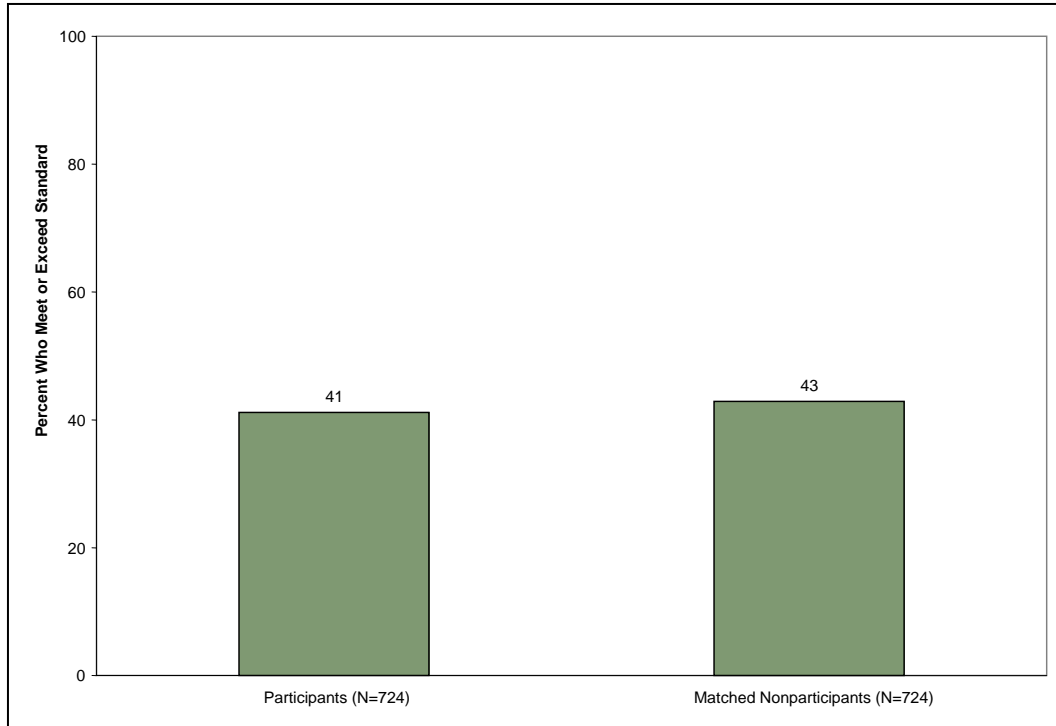


Exhibit reads: Forty-one percent of Cohort 1 participants met or exceeded the state proficiency standard on their tenth-grade AIMS Reading Test.

Cohort 1 Performance on the SRI and the Relationship to Other Achievement Measures

Cohort 1 participants improved their performance on the SRI between the first test point in August 2003 and the last test point in May 2004 (Exhibit 5).⁹ During this period, the average SRI score increased from 21 NCEs (± 1 NCE) to 35 NCEs (± 1 NCE), with this difference in the distribution of scores being statistically significant. The proportion of students scoring below 50 NCEs (below grade level) decreased from 96 percent to 81 percent, and the proportion scoring below 20 NCEs decreased from 45 percent to 18 percent.

⁹ Ninth-graders who completed their first SRI assessment in August 2003 and their last assessment in May 2004 were included in this analysis. Students who scored in the “beginning reader” category at either SRI test point were excluded.

Exhibit 5 Cohort 1 SRI Scores at the Initial and Final Test Points (in NCEs)

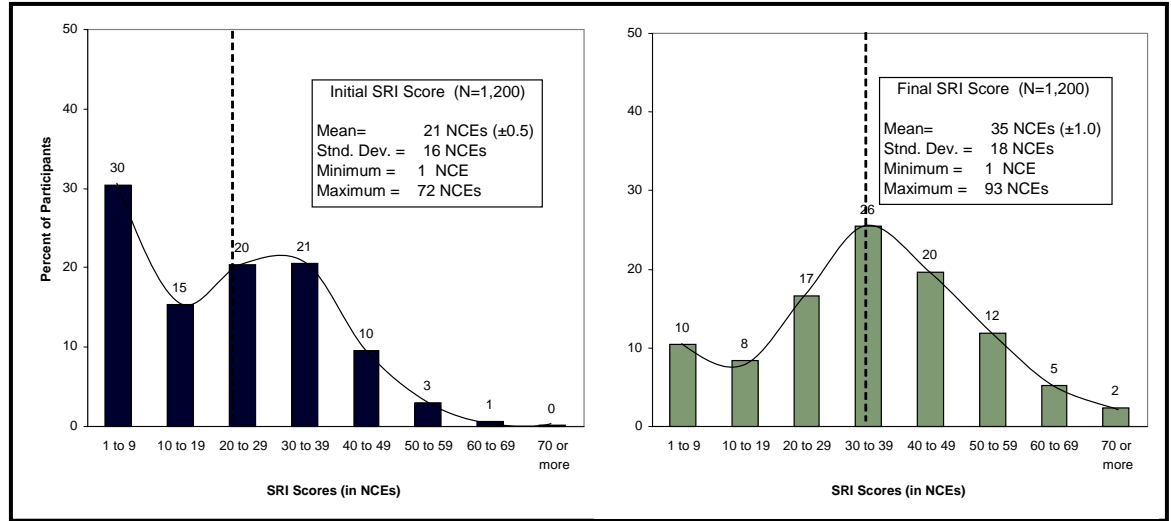


Exhibit reads: The average SRI score at the initial test point was 21 NCEs, compared with an average of 35 NCEs at the final test point.

Exhibit 6 Cohort 1 Changes in SRI Scores (in NCEs)

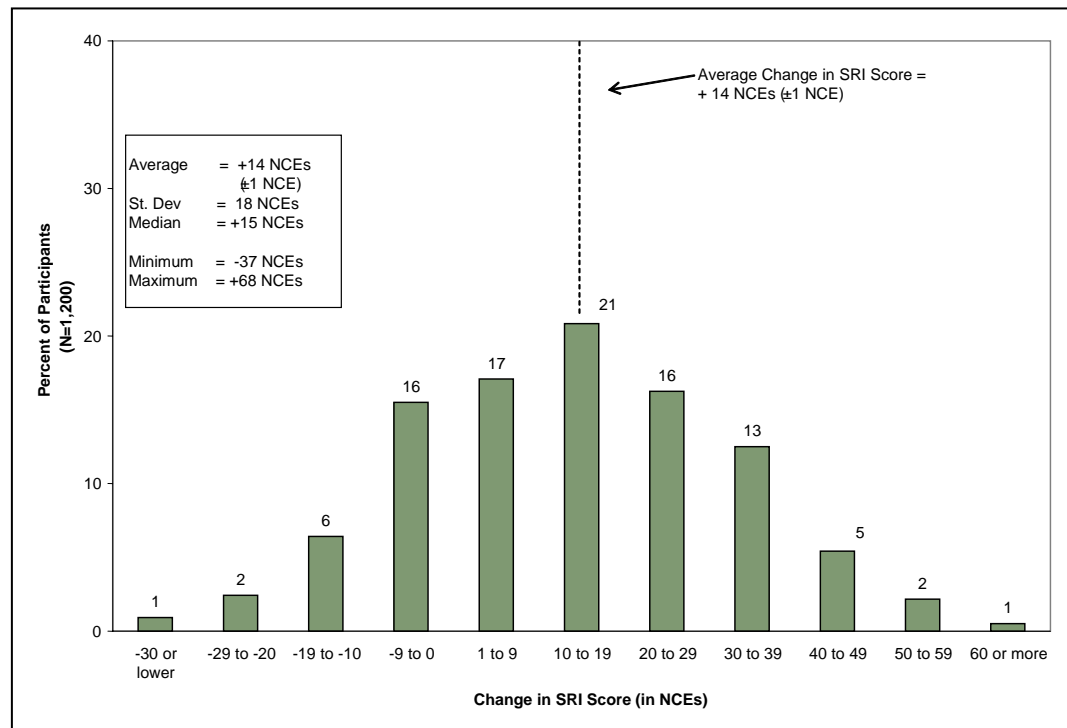


Exhibit reads: The average change in SRI scores between the initial and final test points is a gain of 14 NCEs.

Looking at the individual changes in SRI scores between the first and last test points, these changes ranged from a decrease of 37 NCEs to an increase of 68 NCEs, with an average increase of 14 NCEs (± 1 NCE) (Exhibit 6). Nearly four out of five participants (79 percent) achieved increased SRI scores; scores declined for 21 percent of participants. The differences in the distribution of scores achieved by Cohort 1 participants on the first and last SRI are statistically significant.

In general, it appears that there is a positive but moderate correlation between SRI scores and scores on the SAT9 Reading Comprehension Subtest at the eighth- and ninth-grade test points. The correlation between a student's initial SRI score and his or her eighth grade SAT9 Reading Comprehension Subtest score is positive and statistically significant at 0.33, a statistically significant correlation. Similarly, the correlation between a student's final SRI score and his or her ninth-grade SAT9 Reading Comprehension Subtest is positive and statistically significant at +0.42. There was no relationship between the magnitude of SRI gains among Cohort 1 participants during ninth grade and gains on the SAT9 Reading Comprehension Subtest between eighth grade and ninth grade. The partial correlation, after controlling for their baseline reading proficiency (eighth-grade SAT9 Reading Comprehension Subtest), is +0.06 and is not statistically significant. Together, these correlations could suggest that these two assessments measure different dimensions of reading proficiency.

The analysis did, however, find a relationship between Cohort 1 participants' gains in SRI scores and their scores on the tenth-grade AIMS Reading Test. There is a statistically significant positive partial correlation (0.10) between their AIMS Reading Test scale score and their gains on the SRI, after controlling for their baseline reading proficiency (the eighth-grade SAT9 Reading Comprehension Subtest), indicating that students with larger gains on the SRI were more likely to have higher scores on the AIMS Reading Test. Cohort 1 participants who gained 10 NCEs or more on the SRI averaged 667 scale-score points on the AIMS Reading Test compared with participants who gained less than 10 NCEs and averaged 661 scale-score points. This difference is statistically significant.

Forty-three percent of Cohort 1 participants who gained 10 or more NCEs on the SRI met the performance standards established for the tenth-grade AIMS Reading test, compared with 38 percent of other Cohort 1 participants (Exhibit 7). Seven percent of Cohort 1 participants who gained 10 NCEs or more on the SRI while in READ 180 scored in the "falls far below the standard" performance level on the tenth-grade AIMS Reading test compared with 14 percent of other Cohort 1 participants. The difference in distribution across performance levels between the two groups of participants is statistically significant.

Exhibit 7 Cohort 1 Performance on the Tenth-Grade AIMS Reading Test, By Gains on SRI

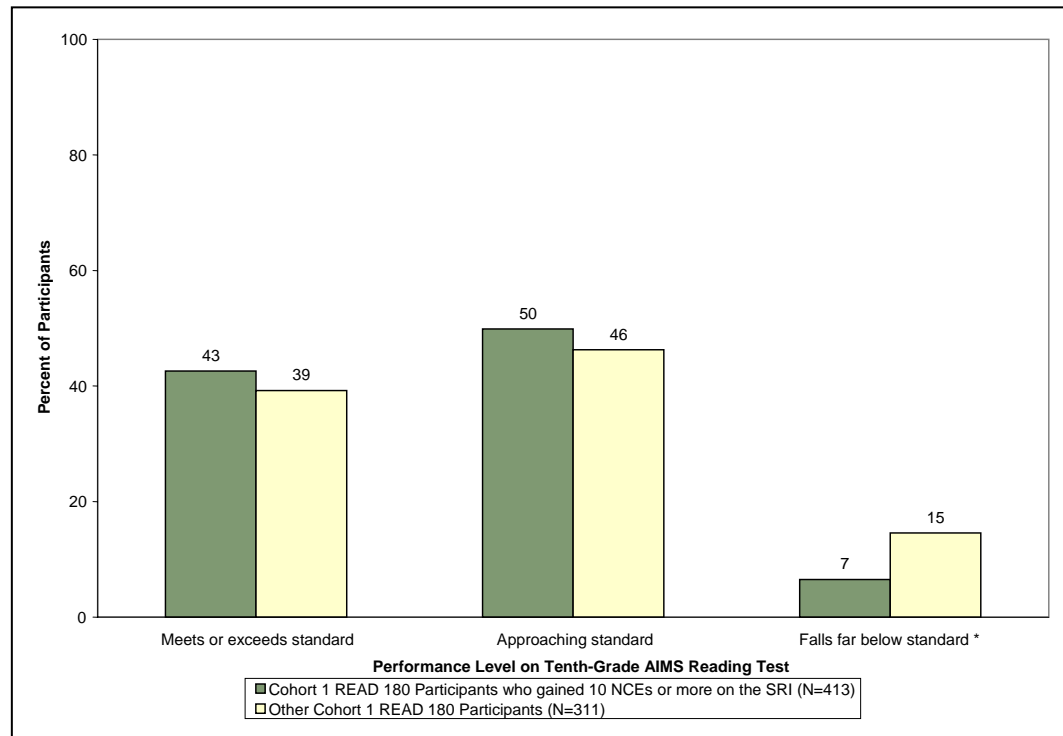


Exhibit reads: Forty-three percent of Cohort 1 participants who gained 10 NCEs on the SRI scored at the “Meets or Exceeds Standard” level on the tenth-grade AIMS Reading Test.

Cohort 1 Subgroups: Performance of ELL Students

Examining the changes in reading proficiency achieved by different subgroups of students can help program managers plan and target services more effectively. In the PUHSD, there is special interest in services to ELL students because of the large numbers of ELL students who attend school in the district. In addition, the variation in reading skills as measured at the eight-grade test point makes it important to examine the relationship between baseline reading scores and gains in reading proficiency to identify options for targeting participation by students who are likely to benefit most from the program.

Cohort 1 ELL READ 180 participants averaged a small gain in reading proficiency as measured by the SAT9 Reading Comprehension Subtest between eighth grade and ninth grade, an *increase* of 0.2 NCEs (Exhibit 8, left-hand panel). Matched nonparticipants eligible for ELL services averaged a *decrease* of 2.8 NCEs. The difference is statistically significant with an effect size of +0.29. This effect size can be interpreted to indicate that 62 percent of the participants scored above the matched participants’ average score. The difference in gains

achieved by Cohort 1 participants and matched nonparticipants who were not eligible for ELL services is not statistically significant.

Exhibit 8
Changes in SAT9 Reading Comprehension Subtest Scores (in NCEs): Cohort 1 ELL Participants and Matched Nonparticipants

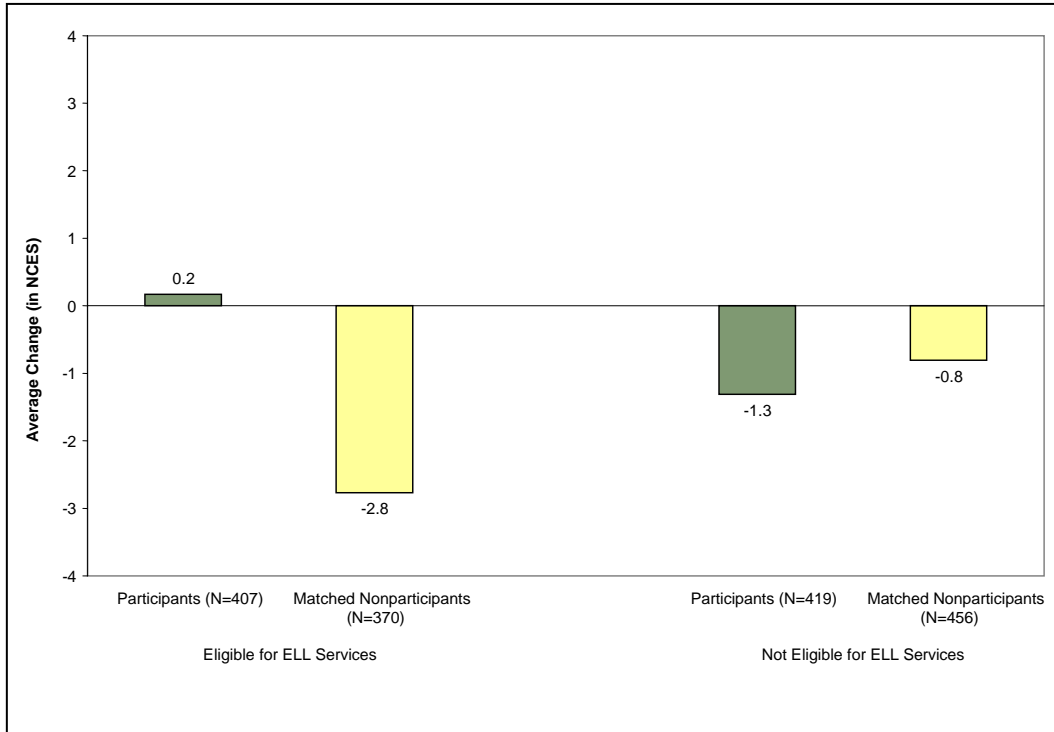


Exhibit reads: Cohort 1 participants eligible for ELL services averaged an increase of 0.2 NCEs between their eighth-grade and ninth-grade SAT9 Reading Comprehension Subtest.

On the AIMS Reading Test, READ 180 ELL participants outperformed matched nonparticipants. On average, these READ 180 participants achieved 654 scale-score points. In contrast, nonparticipating ELL students earned 646 scale-score points. The difference is statistically significant, and the effect size is +0.28. This effect size can also be understood to indicate that 61 percent of the participants scored above the comparison group’s mean score. Correspondingly, a larger proportion of READ 180 ELL participants than matched nonparticipants scored at the “meets or exceeds standard” level in tenth-grade reading, and a larger proportion scored at the “approaching standard” performance level (Exhibit 9).

Exhibit 9
Cohort 1 AIMS Reading Performance Levels:
ELL Participants and Matched Nonparticipants

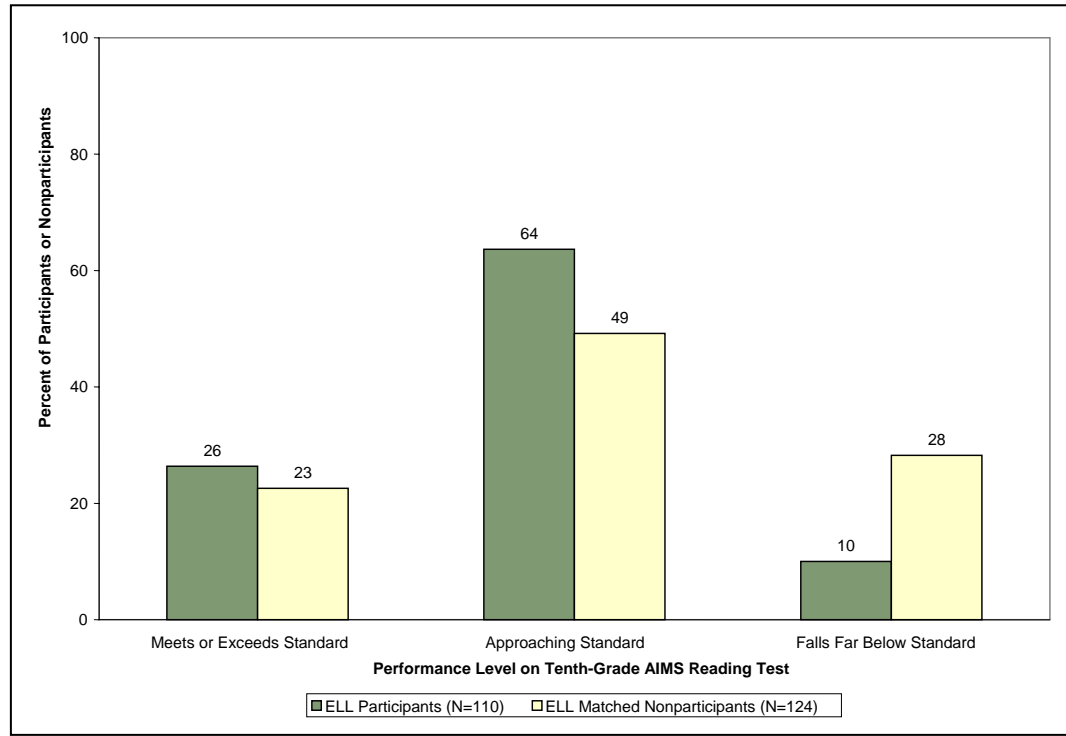


Exhibit reads: Twenty-six percent of ELL students who participated in READ 180 scored at the “meets or exceeds standard” level on the tenth-grade AIMS Reading Test.

Cohort 1 Subgroups: Optimal Range on Measure of Prior Reading Proficiency

Analysis of SAT9 Reading Comprehension Subtest scores of Cohort 1 participants and matched participants found a positive correlation between the scores at the eighth- and ninth-grade test points. Among all of these students, those who had high scores in eighth grade tended to have high scores in ninth grade. The correlation is +0.48, and is statistically significant. However, the correlation between READ 180 participants’ eighth-grade scores and their *change* in scores between eighth grade and ninth grade is negative and statistically significant at -0.39. This negative correlation indicates that students who had lower scores on the eighth grade test tended to achieve larger gains in reading proficiency by the end of ninth grade.

One of the objectives of the analysis of student performance as measured by the state end-of-year assessments for this report is to explore the possibilities of identifying the subgroup of students for whom participation in READ 180 was likely to have the largest benefit, the students scoring in the optimal range on the baseline state assessment. Among Cohort 1 students, it is possible to identify an

upper limit to the optimal range, but there were too few students with scores at the lower end of scores on their eighth-grade SAT9 Reading Comprehension Subtest to allow the identification of a lower score. The upper limit is 45 NCEs.

READ 180 participants who scored 45 NCEs or below on their eighth-grade SAT9 Reading Comprehension Subtest averaged smaller decreases between eighth and ninth grades than did matched nonparticipants (Exhibit 10). These differences are statistically significant, and the effect size is +0.16. This effect size can also be interpreted to indicate that 56 percent of the participants scored above the comparison group's mean score. This pattern was reversed among students who scored above 45 NCEs on their eighth-grade SAT9 Reading Comprehension Subtest, although the difference is not statistically significant. The overall pattern suggests that among Cohort 1 students, those who scored 45 NCEs or below on their eighth-grade SAT9 Reading Comprehension Subtest benefited more from participation in READ 180 than students with higher eighth-grade scores, with the benefit measured as changes in SAT9 scores between eighth grade and ninth grade.

Overall, there is a positive, statistically significant correlation of +0.37 between Cohort 1 participants' scores on the eighth-grade SAT9 Reading Comprehension Subtest and their score on the tenth-grade AIMS Reading Test. A positive correlation indicates that students with higher scores on their eighth-grade test also tended to have higher scores on their tenth-grade test.

Exhibit 10

Cohort 1 Changes in SAT9 Reading Comprehension Subtest Scores (in NCEs): Participants and Matched Nonparticipants

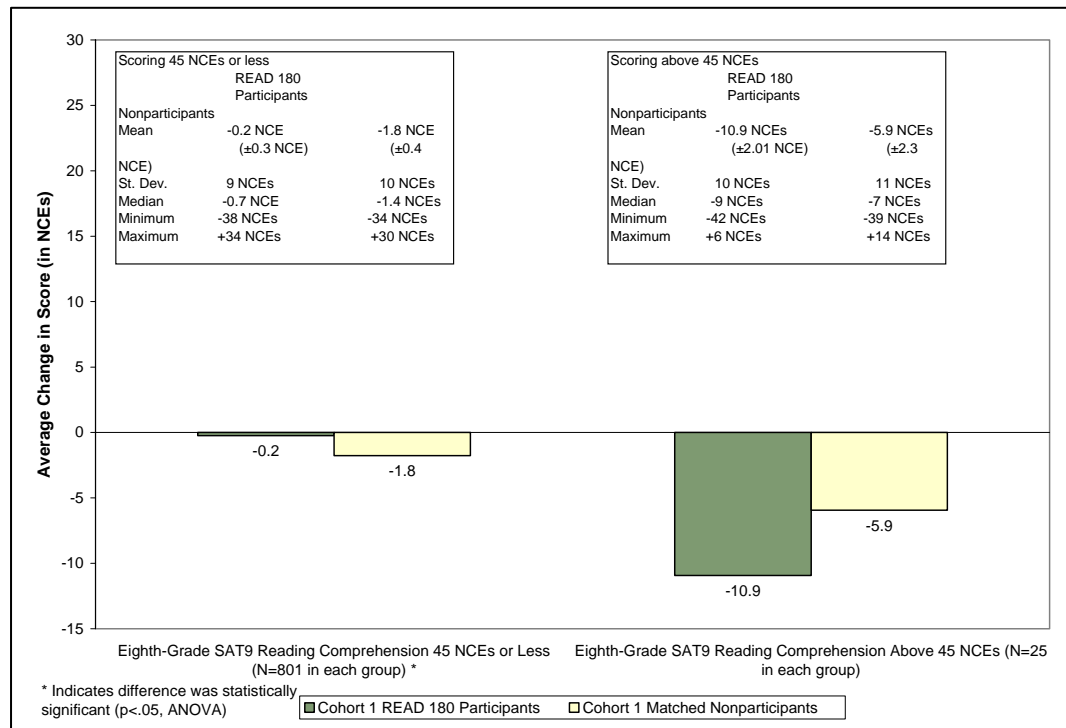


Exhibit reads: Cohort 1 READ 180 participants who scored 45 NCEs or lower on the SAT9 Reading Comprehension Subtest at the eighth-grade test point averaged a decrease of 0.2 NCEs between the eighth- and ninth-grade test points, while the matched nonparticipants averaged a decrease of 1.8 NCEs.

The search for the optimal range of scores on the eighth-grade state assessment associated with the higher scores on the AIMS Reading Test yielded a different threshold for the high end of the range. Comparing Cohort 1 participants to matched nonparticipants, there is a positive relationship between participation in READ 180 and performance on the tenth-grade AIMS Reading Test among students who scored below 35 NCEs on their eighth-grade SAT9 Reading Comprehension Subtest. Cohort 1 participants who scored below 35 NCEs on their eighth-grade SAT9 Reading Comprehension Subtest averaged 656 scale-score points on the AIMS Reading test. Matched nonparticipants averaged 651 scale-score points on the AIMS Reading test. The difference in the distribution of AIMS Reading test scores between the two groups is statistically significant and the effect size is +0.17. An effect size of this magnitude indicates that, among students who scored below 35 NCEs on their eighth-grade test, the relationship between participation in READ 180 and higher scores on the AIMS Reading Test was small but large enough to be substantive. This effect size can also be understood to indicate that 56 percent of the participants scored above the comparison group's mean score.

READ 180 participants who had scored above 35 NCEs on the eighth-grade SAT9 Reading Comprehension Subtest averaged 674 scale-score points on the AIMS Reading Test compared with an average 678 scale-score points among the matched nonparticipants. This difference is not statistically significant.

This pattern suggests that among Cohort 1 students, those who scored 35 NCEs or below on their eighth-grade SAT9 Reading Comprehension Subtest benefited more from participation in READ 180 than did students with higher eighth-grade scores, with the benefit measured as the score on the tenth-grade AIMS Reading Test.

Cohort 2: Students in Ninth Grade in 2004-05

This section of the report examines the changes in reading proficiency among students in Cohort 2, students who were ninth-graders in 2004-05. As with the analysis of changes in reading proficiency achieved by students in Cohort 1, this section presents the results of the analysis of the changes achieved by Cohort 2 participants and matched nonparticipants overall as well the changes achieved by subgroups.¹⁰

Our approach to examining the relationship between participation in READ 180 and reading proficiency for Cohort 2 students differs from the approach we used for examining the relationship among Cohort 1 students. The reason for the different approach is that in spring 2005, the state of Arizona adopted the TerraNova to replace the SAT9. The differences in the constructs measured by the SAT9 Reading Comprehension Subtest and the TerraNova Reading Test and differences in the norming populations preclude the direct computation of the difference in score on the eighth-grade and ninth-grade assessments. The analysis was therefore more limited than the analysis of Cohort 1.

Key findings about the changes in reading proficiency achieved by Cohort 2 students include the following:

- *Cohort 2 participants outperformed matched nonparticipants on the TerraNova Reading Test by a small but statistically significant margin*
- *ELL participants outperformed matched nonparticipants*
- *Participants with lower scores at the eighth-grade test point tended to achieve greater gains in the ninth grade than participants with higher initial scores, and greater gains than matched nonparticipants*

The remainder of this section discusses each of these findings in more detail.

Cohort 2 Participants' Performance on the Spring 2005 TerraNova Reading Test

Both Cohort 2 participants and matched nonparticipants averaged a score of 33 NCEs (± 0.3 NCE) on their eighth-grade SAT9 Reading Comprehension

¹⁰ A subsequent report will examine the performance of Cohort 2 students as tenth-graders, when those data become available.

Subtest (administered in spring 2004). A year later, Cohort 2 participants averaged slightly higher scores than did the matched nonparticipants on the TerraNova Reading Test (Exhibit 11). Specifically, participants averaged 41 NCEs (± 0.3 NCE) and matched nonparticipants averaged 38 NCEs (± 0.4 NCE). The difference is statistically significant, and represents an effect size of +0.27, which in turn can be interpreted to signify that 61 percent of the Cohort 2 participants scored above the matched nonparticipants' average score.

Exhibit 11
Cohort 2 Scores on Eighth-Grade SAT9 Reading Comprehension Subtest and Ninth-Grade TerraNova Reading Test (in NCEs)

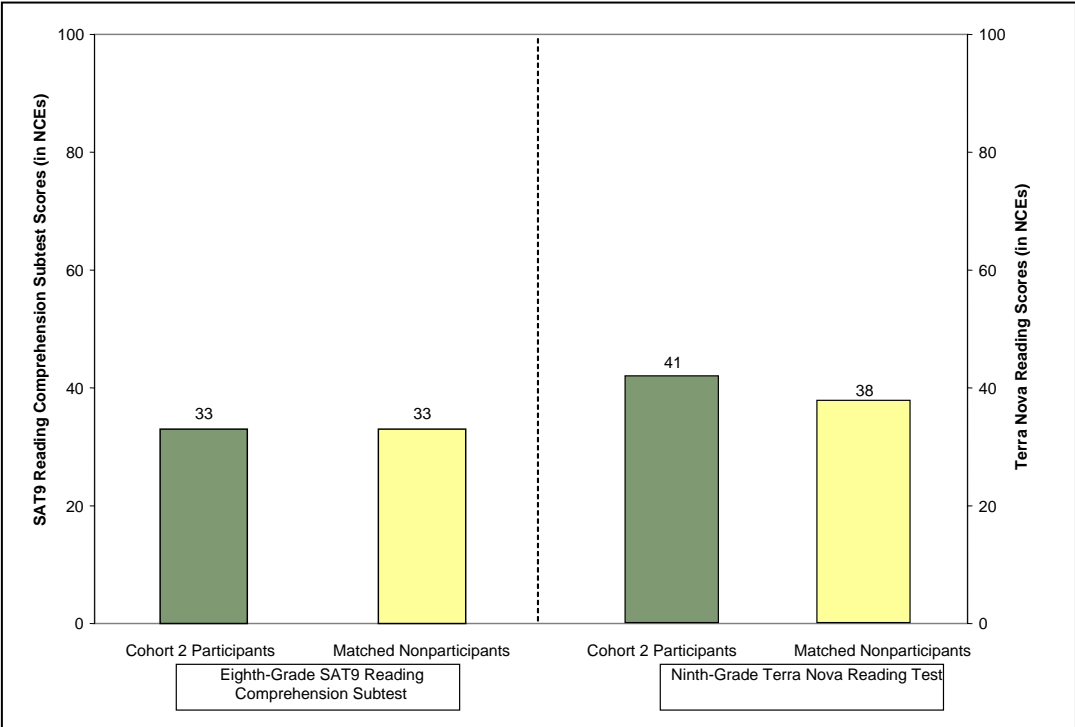


Exhibit reads: Cohort 2 participants and matched nonparticipants averaged a score of 33 NCEs on the eighth-grade SAT9 Reading Comprehension Subtest.

Cohort 2 Performance on the SRI and the Relationship to the TerraNova

Like their Cohort 1 peers, Cohort 2 participants achieved gains on their SRI scores between the first and last test points in 2004-05. Their average SRI score increased from 24 NCEs (± 0.5 NCE) to 34 NCEs (± 0.5 NCE), with the difference in the distribution of scores being statistically significant (Exhibit 12). The proportion of Cohort 2 participants scoring below 50 NCEs decreased from 94 percent to 82 percent, and the proportion scoring below 20 NCEs decreased from 43 percent to 20 percent.

Exhibit 12 Cohort 2 SRI Scores at the Initial and Final Test Points (in NCEs)

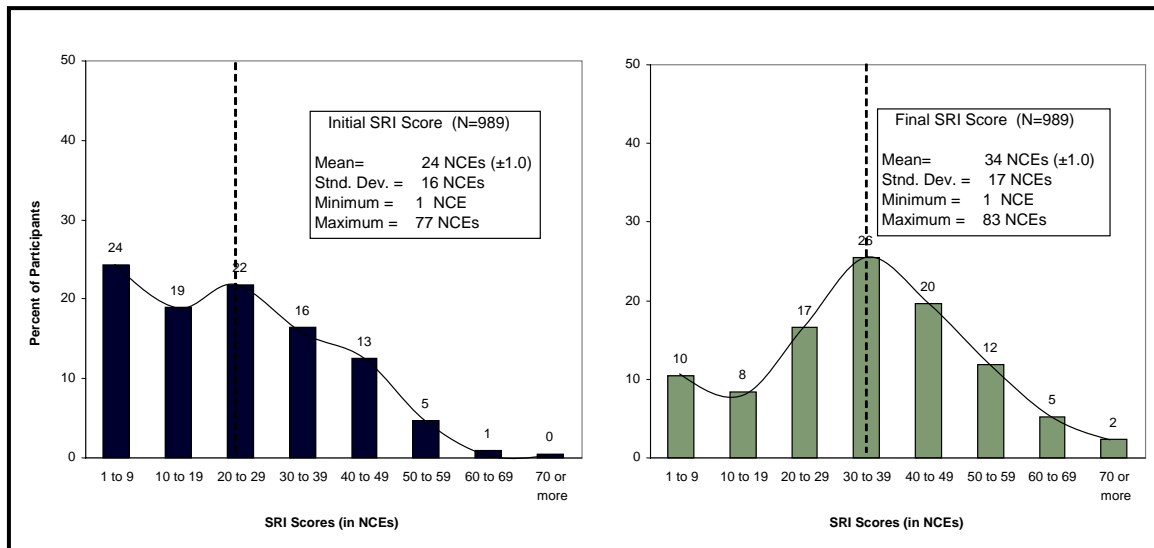


Exhibit reads: The average SRI score at the initial test point was 24 NCEs, compared with an average of 34 NCEs at the final test point.

Among Cohort 2 participants, there is considerable variation in gains in reading proficiency as measured on the SRI, with individual changes ranging from a decrease of 51 NCEs to an increase of 59 NCEs (Exhibit 13). The average change is an increase of 11 NCEs (± 1 NCE). The proportion of participants who increased their SRI scores (72 percent) is greater than the proportion whose scores decrease (25 percent). These differences are statistically significant.

Looking at Cohort 2 participants' performance on the SRI and performance on the TerraNova Reading Test, analysis revealed a positive relationship between gains in reading proficiency as measured on the SRI and scores on the TerraNova Reading test. Overall, the larger a participant's gain on the SRI, the higher that student's score on the TerraNova Reading Test. The correlation between these scores is positive and statistically significant at +0.17.

Participants who gained more than 5 NCEs between their initial and final SRI averaged 42.2 NCEs (± 0.4 NCE) on the 2005 TerraNova Reading test, while those who gained 5 NCEs or less averaged 38.5 NCEs (± 0.5 NCE). This difference is also statistically significant.

Cohort 2 Subgroups: Performance of ELL Students

As with the analysis of the performance of students in Cohort 1, the analysis of the performance of students in Cohort 2 examined the performance of two subgroups—ELL students and students classified according to their baseline reading proficiency. In both cases, analysis revealed variation in subgroup performance.

Exhibit 13
Cohort 2 Changes in SRI Scores

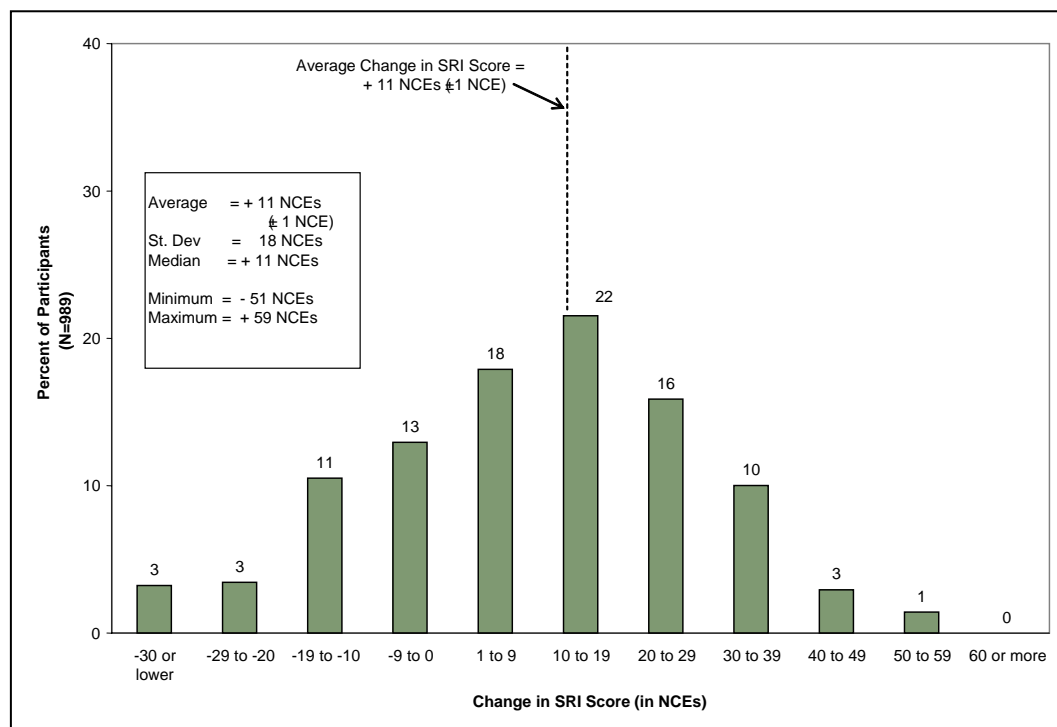


Exhibit reads: The average gain in SRI scores between the initial and final test points was a gain of 14 NCEs.

Cohort 2 ELL participants outperformed matched nonparticipants on the TerraNova Reading Test, with the ELL participants averaging a score of 39.9 NCEs (± 0.4 NCE) on the TerraNova Reading exam and the matched nonparticipants averaging of 35.4 NCEs (± 0.5 NCE). The difference is statistically significant, and the effect size is +0.45, indicating there is a moderate, substantive relationship between participation in READ 180 and performance on the TerraNova among ELL students. This effect size can also be understood to indicate that 67 percent of the ELL participants scored above the matched nonparticipants' average score.

READ 180 participants and nonparticipants who were not eligible for ELL services averaged nearly identical scores on the TerraNova Reading test.

Participants averaged 42.0 NCEs (± 0.4 NCE), while nonparticipants averaged 40.5 NCEs (± 0.5 NCE). The difference between the distribution of scores for the two groups of students is statistically significant, and the effect size is +0.14. This effect size can also be understood to indicate that 55 percent of the participants scored above the matched nonparticipants' average score.

Cohort 2 Subgroups: Optimal Range on Pretest

Cohort 2 participants who scored higher on their eighth-grade SAT9 Reading Comprehension Subtest generally achieved higher scores than did similar nonparticipants. There is a positive correlation between students' scores on their eighth-grade SAT9 Reading Comprehension Subtest and their scores on the TerraNova Reading Test, with the correlation being statistically significant at +0.49, indicating that students with higher scores in eighth grade tended to achieve higher scores in ninth grade.

The analysis of Cohort 2 data extends the exploration of the feasibility of identifying the students who benefit most from participation in READ 180 based upon their scores on the eighth-grade state assessment. Among Cohort 2 participants, the upper threshold of this range was 40 NCEs, but no lower threshold was found. READ 180 participants who had scored 40 NCEs or below on their eighth-grade SAT9 Reading Comprehension Subtest scored higher on the TerraNova Reading Test than did matched nonparticipants with similar scores. Participants averaged a score of 39.8 NCEs on their 2005 TerraNova Reading Test (± 0.3 NCE). Matched nonparticipants averaged 36.2 NCEs (± 0.4 NCEs). The difference between the two groups is statistically significant, and the effect size is +0.30. This effect size can also be understood to indicate that 62 percent of the participants scored above the matched nonparticipants' average score.

Cohort 2 participants and matched nonparticipants who scored above 40 NCEs on their eighth-grade SAT9 Reading Comprehension Subtest averaged nearly the same scores on their ninth-grade TerraNova Reading Test, with participants averaging 46.1 NCEs (± 0.5 NCE) and matched nonparticipants averaging 45.6 NCEs (± 0.5 NCE). The difference between these two groups is not statistically significant.

Cohort 3: Students in Tenth Grade in 2003-04

The analysis of the changes in proficiency achieved by READ 180 participants in Cohorts 3 and 4 is narrower in scope than the analysis of the changes achieved by students in Cohort 1 and 2. This is because the analysis is limited to SRI scores from the initial and final test points. Readers should also note that the PUHSD implementation strategy in the first two years included participation by tenth-graders for only one semester, instead of two semesters.

The key finding from the analysis of Cohort 3 SRI scores is that *the average change in SRI scores was a gain of 7.5 NCEs.*

Cohort 3 Performance on the SRI

Looking at overall changes in SRI scores between the first and last test points, 7 percent of Cohort 3 participants scored 50 NCEs or higher on their initial SRI, and 20 percent did so on their final SRI (Exhibit 14). During this period, the average score increased from 25 NCEs (± 1 NCE) to 32 NCEs (± 1 NCE). These differences are statistically significant.

Exhibit 14
Cohort 3 SRI Scores at the Initial and Final Test Points (in NCEs)

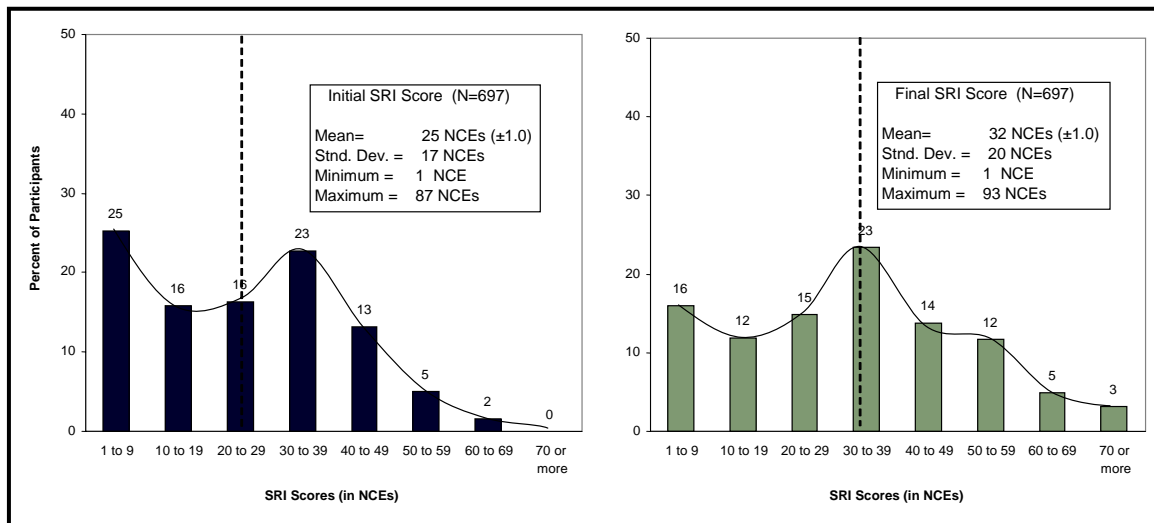


Exhibit reads: The average SRI score at the initial test point was 25 NCEs, compared with an average of 32 NCEs at the final test point.

Not surprisingly, Cohort 3 participants showed smaller changes in SRI scores than did Cohort 1 and Cohort 2 participants. This pattern almost certainly reflects their more limited exposure to the program (one semester versus two

semesters). In addition, one of the properties of the Lexile scale is that it becomes more difficult to increase one’s score at the higher end of the scale than at the lower end. This means that, since tenth-graders typically score higher than ninth-graders, smaller gains would be expected.

Exhibit 15 displays the distribution of changes in individual SRI scores between the first and last test points. Changes in Cohort 3 participants’ SRI scores ranged from a decrease of 42 NCEs to an increase of 61 NCEs. The average change was an increase of 7.5 NCEs (± 1 NCE). Within the overall distribution of changes in SRI scores, just over a third of Cohort 3 participants achieved gains of 15 NCEs or more. At the other end of the spectrum, the scores of 39 percent of the Cohort 3 participants decreased.

Exhibit 15
Cohort 3 Changes in SRI Scores

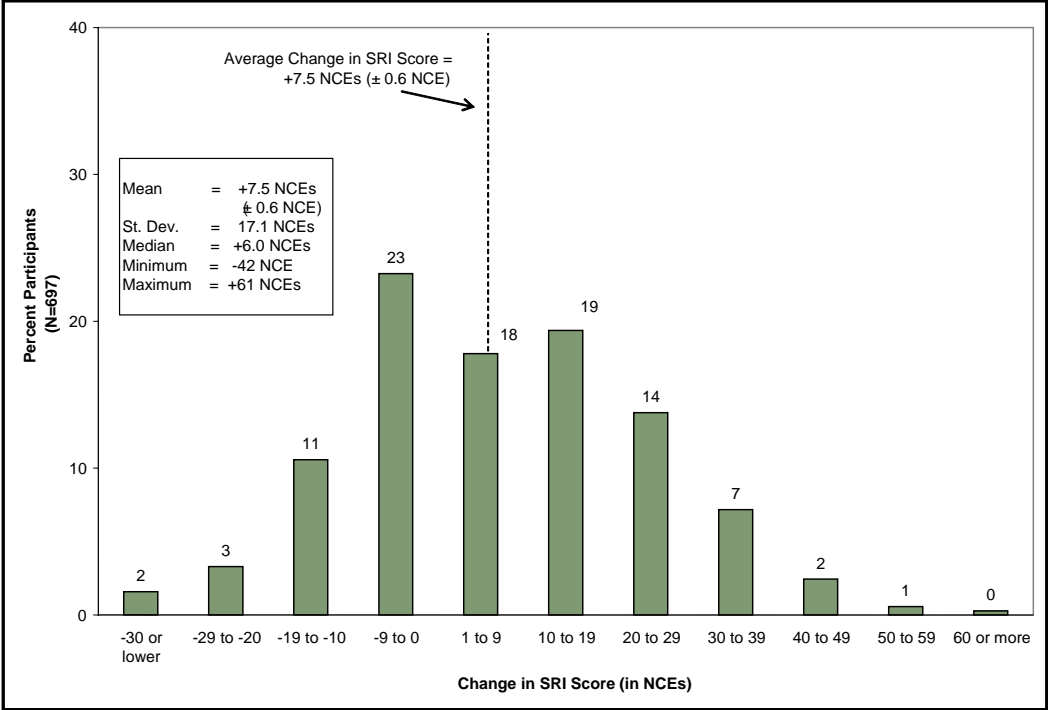


Exhibit reads: The average change in SRI scores between the initial and final test points was 7.5 NCEs.

Cohort 4: Students in Tenth Grade in 2004-05

As noted above, the analysis of changes in reading proficiency achieved by Cohort 4 participants are limited to examining SRI scores. However, Cohort 4 differs from Cohort 3 because Cohort 4 includes two groups of students, which differ in terms of their participation patterns. One group had been enrolled in READ 180 in ninth grade and had been selected to complete an additional semester of READ 180 in tenth grade. These students were also members of Cohort 1 as designated for this study. The second group includes students who were new to READ 180 in 2004-05.

Two findings emerged from our analysis of changes in reading proficiency achieved by Cohort 4 students:

- *Overall, Cohort 4 students achieved only modest increases in reading proficiency.*
- *The small gains on SRI scores achieved by Cohort 4 students who also participated in the program as ninth-graders suggest that the additional semester of participation had little impact on reading proficiency.*

Cohort 4 Performance on the SRI

Cohort 4 students achieved little increase in reading proficiency as measured by the SRI during their semester in the program. The proportion of participants who scored 50 or more NCEs on their initial SRI was 13 percent, increasing to 18 percent on their final SRI (Exhibit 16). The mean score increased from 30 NCEs (± 1 NCE) to 31 NCEs (± 1 NCE). These differences are not statistically significant.

Exhibit 16 Cohort 4 SRI Scores at the Initial and Final Test Points (in NCEs)

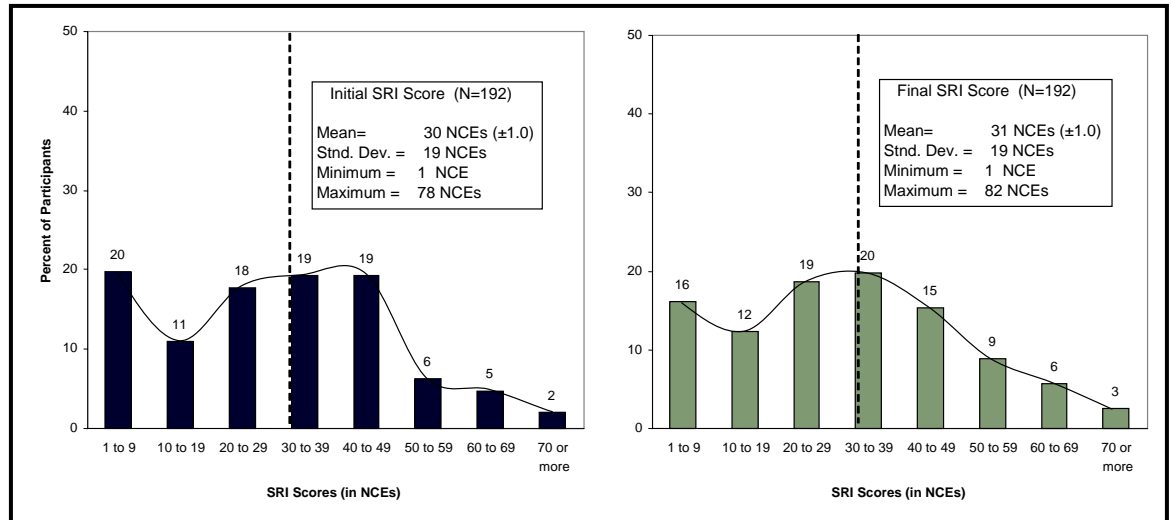


Exhibit reads: The average SRI score at the initial test point was 30 NCEs, compared with an average of 31 NCEs at the final test point.

Changes in individual SRI scores during the semester ranged from a decrease of 42 NCEs to an increase of 43 NCEs, with an average increase of 1.5 NCEs (± 1 NCE) (Exhibit 17). Overall, the proportion of Cohort 4 students who achieved a gain in the SRI scores was virtually the same as the proportion whose scores decreased. In general, participants with lower scores on their initial SRI achieved larger gains on the SRI than did other participants. The correlation between initial SRI scores and the change in SRI scores is -0.37 , a statistically significant correlation.

Exhibit 17 Cohort 4 Changes in SRI Scores

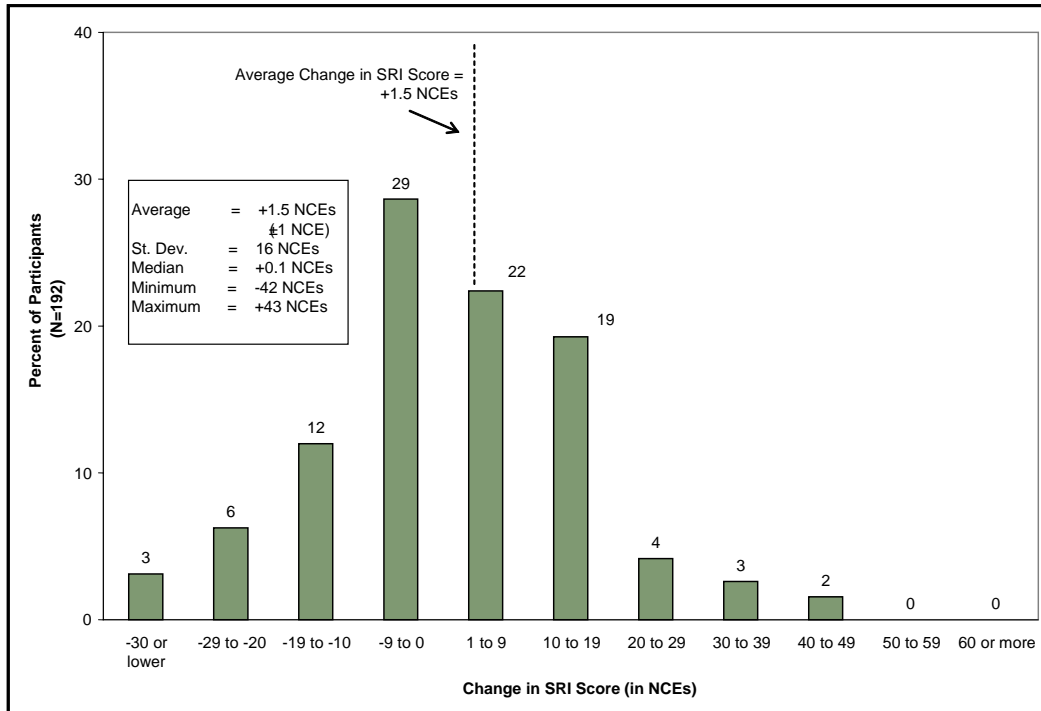


Exhibit reads: The average change in SRI scores between the initial and final test points was a gain of 1.5 NCEs.

Cohort 4 READ 180 Participants Participating for an Additional Semester

Overall, the two groups in this cohort were similar in reading ability at the start of tenth grade. Both groups averaged an initial SRI score of 30 NCEs (Exhibit 18).

Exhibit 18
Cohort 4 Initial SRI Scores (in NCEs):
New and Continuing Participants

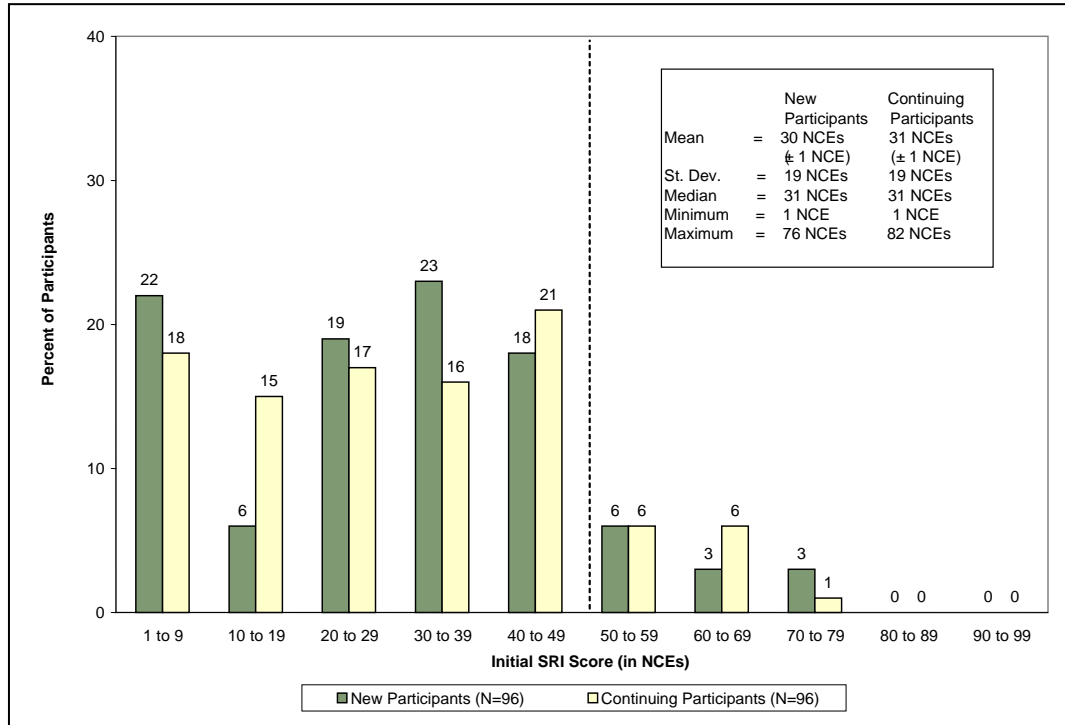


Exhibit reads: Among new Cohort 4 participants the average SRI score at the initial test point was 30 NCEs, compared with an average SRI score of 31 NCEs among the continuing participants.

Both the new and the continuing READ 180 participants in Cohort 4 achieved small gains between their initial and final SRI. New participants averaged a gain of 1.9 NCEs, while continuing participants averaged a gain of 1.0 NCE (Exhibit 19). The difference between the two groups is not statistically significant.

Exhibit 19
Cohort 4 Changes in SRI Scores (in NCEs):
New and Continuing Participants

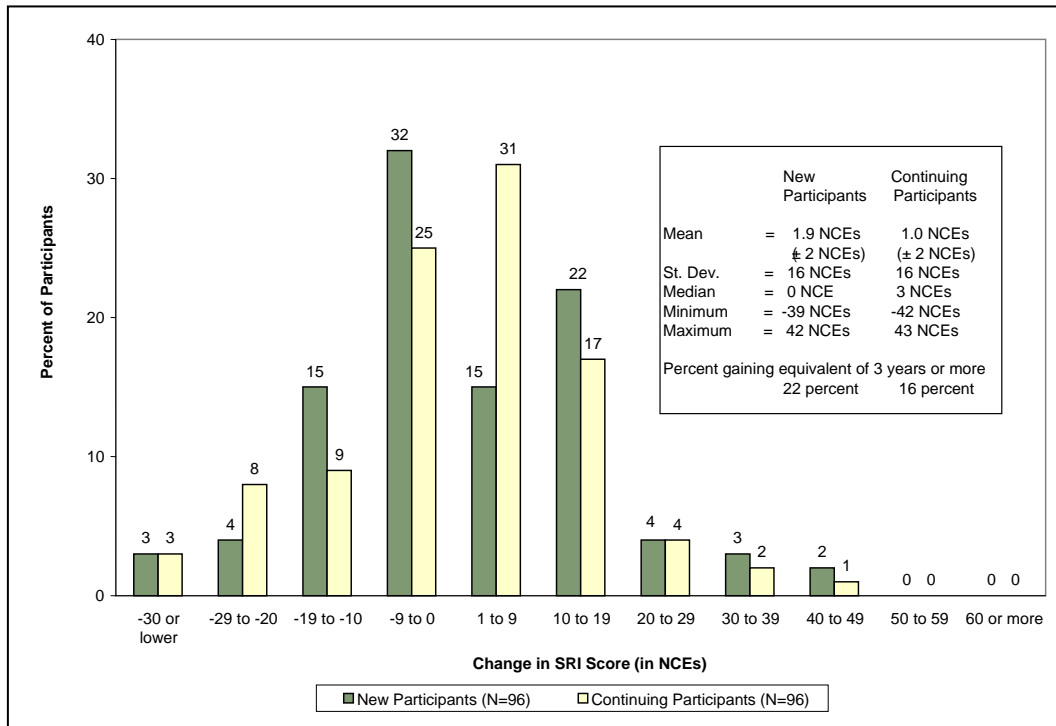


Exhibit reads: Among new Cohort 4 participants, the average change in SRI scores between the initial and final test points was a gain of 1.9 NCEs compared with an average change of a gain of 1 NCE among continuing participants.

Participants who had been in both Cohorts 1 and 4 had averaged a gain of 14.2 NCEs on their ninth-grade SRI, slightly greater than the gain of 13.7 NCEs achieved by participants who were only in Cohort 1 (Exhibit 20). The relatively small gains achieved by these students in their additional tenth-grade semester, compared with the large gains achieved in ninth grade, suggest that the additional period of participation was not associated with continued gains in reading proficiency.

Exhibit 20 Cohort 4 SRI Scores (in NCEs) New and Continuing Participants

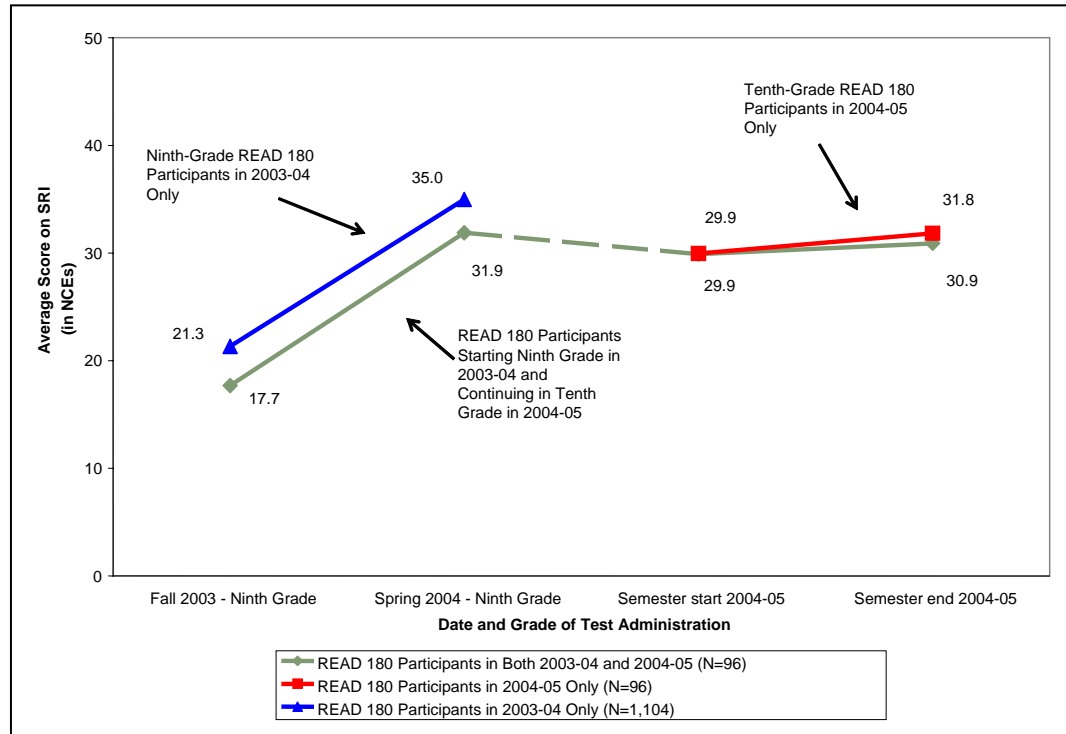


Exhibit reads: Cohort 1 participants who participated as ninth-graders only in 2003-04 averaged an initial SRI score of 21.3 NCEs, compared with 17.7 NCEs among the Cohort 1 participants who would continue to participate during tenth grade.

Conclusions and Suggestions for Additional Research

This study of four cohorts of PUHSD students in 2003-04 and 2004-05 found that participation in READ 180 was associated with meaningful benefits in reading proficiency.

The performance of 2003-04 READ 180 participants on the tests administered as part of the Arizona state assessment showed benefits in terms of gains in reading proficiency. The SAT9 Reading Comprehension Subtest scores of Cohort 1 participants decreased less between eighth and ninth grades than did the scores of a group of similar nonparticipants who were matched on the bases of eighth-grade test scores and demographic characteristics. The differences between the participants and matched nonparticipants are statistically significant. However, there are no overall differences between Cohort 1 participants and matched nonparticipants on the tenth-grade AIMS Reading Test.

Analysis of trends in the reading proficiency of Cohort 2 participants was complicated by the change in the standardized test that was administered. These students completed the SAT9 Reading Comprehension Subtest in eighth grade and the TerraNova Reading test in ninth grade. READ 180 participants averaged higher scores on the ninth-grade TerraNova Reading Test than did the matched group of nonparticipants.

Among Cohort 1 and Cohort 2 students, two subgroups of students benefited more from participation in READ 180. READ 180 participants who were eligible for ELL services achieved larger gains on the state assessments than did the matched nonparticipants who were eligible for ELL. These differences were observed in changes on the state assessments between eighth grade and ninth grade, and again in the differences in the average scores achieved by participants and matched nonparticipants on the tenth-grade reading tests.

The second group of students who benefited more from READ 180 were students who had lower levels of reading proficiency prior to participating in the program. For example, 2004-05 ninth-grade READ 180 participants who had scored 40 NCEs or below on their eighth-grade SAT9 Reading Comprehension Subtest scored higher on their ninth-grade TerraNova Reading Test than did the matched group of nonparticipants.

Tenth-grade students who participated in READ 180 for one semester in 2003-04 averaged SRI gains of more than 5 NCEs. Gains among the 2004-05 cohort of tenth-grade participants were smaller.

The findings from this study also suggest several areas for additional research and analysis to improve understanding of the relationship between participation in READ 180 and academic performance.

- ***A broader range of academic outcomes associated with participation in READ 180.*** In addition to changes in reading proficiency as measured by the SRI, TerraNova Reading Test, and AIMS Reading Test, it would be useful to explore whether there is any relationship with indicators of attachment to school and progress toward high school graduation, including school attendance, suspensions/expulsions, course grades, course credits earned, and grade promotion/retention.
- ***The relationship between the quality of implementation and sustained use of READ 180 and changes in reading proficiency.*** Examining the extent to which varying amounts of gains in student reading proficiency are associated with key features of program implementation, including implementation of the program's complex instructional model, would help identify needs for enhanced implementation and teacher professional development.
- ***The collection and analysis of data for additional years and cohorts of students.*** The comparison of the performance of the Cohort 1 READ 180 participants and matched participants on their eleventh-grade AIMS tests and other areas of academic performance would yield additional information on the persistence of the relationship between participation in READ 180 and academic performance. Analysis of the performance of the Cohort 2 READ 180 participants and matched nonparticipants on their tenth-grade AIMS Reading Tests and other areas of academic performance would strengthen the findings about the Cohort 1 students presented in this report. Analysis of the results for the ninth-grade students enrolled in READ 180 during 2005-06 would allow the measure of change in reading proficiency between eighth and ninth grades on the TerraNova Reading Test. This would allow the exploration of whether the relationship between participation and gains on the SAT9 Reading Comprehension Subtest were also to be found on the TerraNova Reading Test.

Appendix A

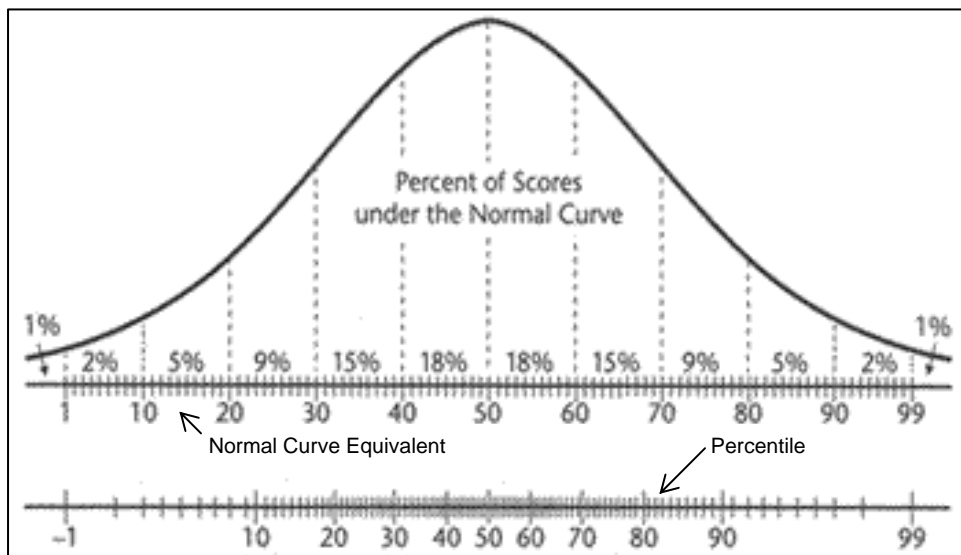
Using Normal Curve Equivalent Scores

A normal curve equivalent score (NCE) is a standardized score (based on a normal distribution) that makes it possible to compare scores across grades and to compute statistics that describe the performance of a group of students, such as average scores, that are not possible using other ways of expressing test scores, such as percentiles. NCEs consist of 99 equal units, an advantage over percentiles and grade equivalents in analyzing changes in test scores. Because NCE scores are equal units, they can be averaged for a group of students.

NCEs are computed by dividing the normal curve distribution of student scores on an assessment into 99 units, with each NCE unit spanning the same number of test points. NCE scores have a mean of 50 and a standard deviation of 21.06. In a normal distribution, an NCE of 1.0 is equivalent to a percentile rank of 1.0, and an NCE of 99.0 is equivalent to a percentile rank of 99.0.

Percentile ranks, as can be seen in Exhibit A1, tend to cluster around the middle of the distribution of test scores, so that a change in performance of one percentile represents a smaller change in the raw score on an assessment at the center of the distribution than at the extremes of the distribution. Because NCEs comprise 99 units of equal size, a change of one NCE represents the same change in the raw score at any point in the distribution.

Exhibit A-1
Distribution of Student Scores, Normal Curve Equivalent Scores,
and Percentiles



Students performing at grade level will have an NCE of 50, regardless of their grade level. If a student makes exactly one year of progress after one year of instruction, his/her NCE score would be exactly the same both years, and the change in NCE scores would be zero. A student with a one-year change in NCE scores that is greater than zero has achieved larger gains than the general population, and a student with a negative change in NCE scores has made less progress.

Appendix B

Propensity Matching Procedures for Identifying Matched Comparison Students

We used propensity matching to identify sets of comparison students from the full population of non-participating PUHSD students. The comparison students were selected to be as close as possible to the READ 180 students with respect to reading proficiency as measured on the eighth-grade SAT9 Reading Comprehension Subtest. Among Cohort 1 students, eligibility for English Language Learner (ELL) services and for special education services was included in the matching procedure. More data were available for Cohort 2 students. Among these students, matching was conducted using eighth-grade reading proficiency, ELL and special education eligibility, gender, and ethnicity.

Propensity matching is an algorithm available for use in the SAS statistical software package. In propensity matching, we start with a pool of potential comparison cases and select one-to-one matches for each student such that we minimize the “distance” between READ 180 participants and the selected comparison students. “Distance” refers to the overall difference in group means (here measured in standard deviations) on each matching variable, which varied by cohort of students because of the differences in the data available. The automated matching procedure repeats itself, reorganizing and replacing potential matches, until the overall distance between the two groups is minimized. Thus, in the first round, all identified matches are potentially temporary and may be replaced if later rounds identify a different set of matches that reduces the distance between the two groups. Propensity matching is a process separate from outcome analyses. It simply identifies the cases to be used in subsequent analyses and in no other way affects those analyses.

Exhibit B-1 displays the variables used to match students, the mean values for READ 180 participants and the matched nonparticipants, and the average distance between the two groups, measured in standard deviations.

Exhibit B-1
Characteristics of READ 180 Participants and Matched Nonparticipants

	Participants	Matched Nonparticipants
<i>Cohort 1</i>		
	<i>Mean score:</i>	
Eighth-grade achievement (SAT9 Reading Comprehension Subtest)	32.0 NCEs	32.0 NCEs
	<i>Mean percent of students:</i>	
ELL eligibility status	51 Percent	55 Percent
Special Education eligibility status	8 percent	8 percent
		<i>Mean difference, in standard deviations:</i>
Average Distance		-0.03
<i>Cohort 2</i>		
	<i>Mean score:</i>	
Eight-grade achievement	33.0	33.0
	<i>Mean percent of students:</i>	
ELL eligibility status	40 percent	44 percent
Special Education eligibility status	7 percent	10 percent
Percent Female	48 percent	49 percent
Percent Hispanic	84 percent	86 percent
		<i>Mean difference, in standard deviations:</i>
Average Distance		-0.03

Appendix C

Central Tendencies and Distributions

Exhibit C-1 Cohort 1 Reading Achievement Outcomes (in NCEs), by Measure and Subgroup

<i>Measure and Group</i>	Mean	Median	SD	Range
<i>Cohort 1</i>				
<i>Eighth-Grade SAT9 Reading Comprehension Subtest</i>				
Participants (N=826)	32.0	33.7	8.9	1 - 72
Matched Nonparticipants (N=826)	32.0	33.7	8.9	1 - 72
<i>Ninth-Grade SAT9 Reading Comprehension Subtest</i>				
Participants (n=826)	31.4	32.3	9.3	1 - 64
Matched Nonparticipants (N=826)	30.1	31.5	11.3	1 - 64
Participants eligible for ELL services (N=419)	32.4	32.3	8.8	1 - 64
Matched Nonparticipants eligible for ELL services (N=456)	28.7	29.9	11.7	1 - 64
Participants scoring 45 NCEs or below on eighth-grade test (N=801)	31.2	32.3	9.2	1 - 64
Matched Nonparticipants scoring 45 NCEs or below eighth-grade test (N=801)	29.6	31.5	11.0	1 - 64
Participants scoring above 45 NCEs on eighth-grade test (N=25)	40.3	39.6	8.4	26 - 62
Matched Nonparticipants above 45 NCEs on eighth-grade test (N=25)	45.3	43.0	8.6	30 - 64
<i>Tenth-Grade AIMS Reading (Scale Scores)¹¹</i>				
Participants (N=724)	664.1	667.0	28.5	581 - 745
Matched Nonparticipants (N=724)	664.2	667.0	31.9	550 - 805
Participants eligible for ELL services (N=110)	653.7	652.0	26.2	581 - 720
Matched Nonparticipants eligible for ELL services (N=124)	645.6	638.0	29.0	550 - 710
Participants scoring 35 NCEs or below on eighth-grade test (N=408)	656	656.0	27.4	581 - 731
Matched Nonparticipants scoring 35 NCEs or below eighth-grade test (N=374)	651	652.0	29.4	550 - 731
Participants scoring above 35 NCEs on eighth-grade test (N=316)	674	674.0	26.7	597 - 745
Matched Nonparticipants 35 NCEs on eighth-grade test (N=351)	678	677.0	28.4	574 - 805

¹¹ The results of the AIMS Reading Test were not available in NCEs.

Measure and Group	Mean	Median	SD	Range
SRI, Fall 2003				
Participants (N=1,200)	21.0	22.0	16.0	1 - 72
Participants eligible for ELL services (N=431)	21.8	23.0	15.5	1 - 60
Participants scoring 45 NCEs or below on eighth-grade test (N=857)	20.9	22.0	15.6	1 - 67
Participants scoring above 45 on eighth-grade test (N=25)	35.8	36.0	18.1	1 - 66
SRI, Spring 2004 (Lexile Scale Scores)				
Participants (N=1,200)	34.7	36.0	17.8	1 - 93
Participants eligible for ELL services (N=431)	35.1	36.0	17.4	1 - 87
Participants scoring 45 NCEs or below on eighth-grade test (N=857)	34.2	35.0	17.4	1 - 93
Participants scoring above 45 on eighth-grade test (N=25)	53.2	55.0	18.9	1 - 87

Exhibit C-2
Cohort 2 Reading Achievement Outcomes (in NCEs),
by Measure and Subgroup

<i>Measure and Group</i>	Mean	Median	SD	Range
<i>Spring 2004 Eighth-Grade SAT9 Reading Comprehension Subtest</i>				
Participants (N=815)	33.0	33.7	8.9	1 - 62
Matched nonparticipants (N=815)	33.0	33.7	8.9	1 - 62
<i>Spring 2005 TerraNova Reading Test</i>				
Participants (N=815)	41.2	42.0	8.9	1 - 70
Matched nonparticipants (N=815)	38.3	40.0	12.2	1 - 77
Participants eligible for ELL services (N=325)	39.9	41.0	7.7	15 - 66
Matched nonparticipants eligible for ELL (N=361)	35.4	37.0	11.1	1 - 69
Participants not eligible for ELL services (N=490)	42.0	43.0	9.4	1 - 70
Matched nonparticipants not eligible for ELL (N=454)	40.5	43.0	12.7	1 - 77
Participants scoring 40 NCEs or below on eighth-grade test (N=634)	39.8	41.0	8.4	1 - 67
Matched nonparticipants scoring 40 NCEs or below eighth-grade test (N=634)	36.2	38.0	12.2	1 - 66
Participants scoring above 40 NCEs on eighth-grade test (N=181)	46.1	46.0	8.4	9 - 70
Matched nonparticipants above 40 NCEs on eighth-grade test (N=181)	45.6	47.0	9.2	6 - 77
Participants who gained more than 5 NCEs on SRI (N=307)	42.2	43.0	8.3	1 - 70
Participants who gained 5 NCEs or less on SRI (N=508)	39.5	40.0	9.4	1 - 66
<i>SRI, Fall 2004</i>				
Participants (N=815)	23.5	23.0	16.3	1 - 77
Participants eligible for ELL services (N=325)	19.2	17.8	15.6	1 - 76
Participants not eligible for ELL services (N=490)	26.3	27.1	16.2	1 - 77
Participants scoring 40 NCEs or below on eighth-grade test (N=634)	21.9	20.9	15.9	1 - 76
Participants scoring above 40 NCEs on eighth-grade test (N=181)	28.8	28.7	16.7	1 - 77
<i>SRI, Spring 2005</i>				
Participants (N=815)	34.3	33.8	17.3	1 - 83
Participants eligible for ELL services (N=325)	30.4	30.7	15.8	1 - 76
Participants not eligible for ELL services (N=490)	37.0	27.1	16.2	1 - 77
Participants scoring 40 NCEs or below on eighth-grade test (N=634)	31.9	31.8	16.8	1 - 83
Participants scoring above 40 NCEs on eighth-grade test (N=181)	42.9	42.6	16.2	1 - 80

Exhibit C-3
Cohort 3 SRI Reading Achievement Outcomes (in NCEs)

<i>Measure and Group</i>	Mean	Median	SD	Range
<i>SRI, Fall 2003</i> Participants (N=697)	24.6	26.0	17.1	1 - 87
<i>SRI, Spring 2004</i> Participants (N=697)	32.0	32.0	19.7	1 - 93

Exhibit C-4
Cohort 4 SRI Reading Achievement Outcomes (in NCEs)

<i>Measure and Group</i>	Mean	Median	SD	Range
<i>SRI, Fall 2004</i> Participants (N=192)	29.9	30.6	19.0	1 - 78
<i>SRI, Spring 2005</i> Participants (N=192)	31.4	31.4	19.3	1 - 82

Appendix D

Crosswalk between SRI Lexile Scale Scores and Normal Curve Equivalent Scores

Grade 9				Grade 10			
SRI Lexile Scale Score Range	NCE	SRI Lexile Scale Score Range	NCE	SRI Lexile Scale Score Range	NCE	SRI Lexile Scale Score Range	NCE
Below 585	1	1,050 to 1,059	51	Below 629	1	1,085 to 1,094	51
585 to 619	7	1,060 to 1,069	52	630 to 664	7	1,095 to 1,104	52
620 to 644	10	1,070 to 1,084	53	665 to 689	10	1,105 to 1,114	53
645 to 664	13	1,085 to 1,094	54	690 to 709	13	1,115 to 1,124	54
665 to 679	15	1,095 to 1,104	55	710 to 729	15	1,125 to 1,134	55
680 to 694	17	1,105 to 1,114	56	730 to 744	17	1,135 to 1,144	56
695 to 709	19	1,115 to 1,119	57	745 to 759	19	1,145 to 1,149	57
710 to 724	20	1,120 to 1,129	58	760 to 769	20	1,150 to 1,159	58
725 to 734	22	1,130 to 1,139	59	770 to 779	22	1,160 to 1,169	59
735 to 744	23	1,140 to 1,149	60	780 to 789	23	1,170 to 1,179	60
745 to 754	24	1,150 to 1,154	61	790 to 799	24	1,180 to 1,184	61
755 to 764	25	1,155 to 1,164	62	800 to 809	25	1,185 to 1,194	62
765 to 774	26	1,165 to 1,169	63	810 to 819	26	1,195 to 1,199	63
775 to 784	27	1,170 to 1,184	64	820 to 829	27	1,200 to 1,209	64
785 to 794	28	1,185 to 1,189	65	830 to 839	28	1,210 to 1,214	65
795 to 804	29	1,190 to 1,204	66	840 to 849	29	1,215 to 1,229	66
805 to 814	30	1,205 to 1,209	67	850 to 859	30	1,230 to 1,234	67
815 to 824	31	1,210 to 1,224	68	860 to 869	31	1,235 to 1,249	68
825 to 844	32	1,225 to 1,229	69	870 to 889	32	1,250 to 1,259	69
845 to 854	33	1,230 to 1,239	70	890 to 904	33	1,260 to 1,269	70
855 to 869	34	1,240 to 1,244	71	905 to 919	34	1,270 to 1,274	71
870 to 879	35	1,245 to 1,259	72	920 to 929	35	1,275 to 1,284	72
880 to 899	36	1,260 to 1,264	73	930 to 944	36	1,285 to 1,289	73
900 to 904	37	1,265 to 1,274	74	945 to 849	37	1,290 to 1,299	74
905 to 919	38	1,275 to 1,279	75	950 to 964	38	1,300 to 1,304	75
920 to 929	39	1,280 to 1,289	76	965 to 969	39	1,305 to 1,314	76
930 to 944	40	1,290 to 1,299	77	970 to 984	40	1,315 to 1,324	77
945 to 959	41	1,300 to 1,309	78	985 to 994	41	1,325 to 1,334	78
960 to 974	42	1,310 to 1,324	80	995 to 1009	42	1,335 to 1,349	80
975 to 979	43	1,325 to 1,339	81	1,010 to 1,014	43	1,350 to 1,364	81
980 to 994	44	1,340 to 1,359	83	1,015 to 1,024	44	1,365 to 1,384	83
995 to 1,004	45	1,360 to 1,379	85	1,025 to 1,034	45	1,385 to 1,404	85
1,005 to 1,014	46	1,380 to 1,409	87	1,035 to 1,044	46	1,405 to 1,439	87
1,015 to 1,024	47	1,410 to 1,454	90	1,045 to 1,054	47	1,440 to 1,479	90
1,025 to 1,034	48	1,455 to 1,504	93	1,055 to 1,064	48	1,480 to 1,529	93
1,035 to 1,044	49	1,505 and higher	99	1,065 to 1,079	49	1,530 and higher	99
1,045 to 1,049	50	,		1,080 to 1,084	50	,	