THE UNIVERSITY OF MEMPHIS | The Center for Research in Educational Policy (CREP)

The LASER Model: A Systemic and Sustainable Approach for Achieving High Standards in Science Education

Summative Report Section 6: State Assessments

Yun Tang, B.A.

Todd Zoblotsky, Ph.D.

The University of Memphis

7/15/2015

Acknowledgments

The success of this evaluation would not have been possible without the herculean efforts built on strong partnerships among the Center for Research in Educational Policy (CREP), the Smithsonian Science Education Center (SSEC), Abt Associates, Bernalillo Public Schools, Chama Public Schools, Cleveland County Schools, Greene County Schools, Houston Independent School District, Jemez Valley Public Schools, Johnston County Schools, Los Alamos Public Schools, McDowell County Schools, Moore County Schools, Mora Public Schools, Pecos Independent School District, Rio Rancho Public Schools, Santa Fe Public Schools, Warren County Schools, and Wilson County Schools. We extend our heartfelt thanks and appreciation to all who contributed to this amazing endeavor, and sought – and still seek – to improve the state of science education in America.

CREP Project Staff:

Marty Alberg	Principal Investigator
Carolyn Kaldon	Co-Principal Investigator
Dan Strahl	Co-Principal Investigator
Michael Rowe	Project Manager
John Burgette	Qualitative Analysis
Todd Zoblotsky	Statistics
Lou Franceschini	Statistics
Haixia Qian	Statistics
Bryan Winter	Statistics
Ying Huang	Statistics
Adrian Young	School Liaison
Cindy Muzzi	School Liaison
Dallas Burkhardt	Site Researcher Liaison
Margie Stevens	SMS Administration
Ruby Booth	SMS Administration

Introduction

State assessment results from the Spring 2014 (third posttest for the Elementary and the Middle School Cohort) administrations are currently available and are reported below. As the state assessments for each region are different, results are reported by region only (and not in the aggregate across regions).

Outcomes for the Houston Independent School District (HISD) are reported first (State of Texas Assessment of Academic Readiness (STAAR) and Stanford), followed by the New Mexico (Standards Based Assessment (SBA)) and North Carolina (End-of-Grade (EOG)) regions. A summary of the Key Findings for each set of analyses is presented at the beginning of each report, followed by information on the samples included, baseline equivalence between the Phase 1 and Phase 2 groups, and the detailed outcomes by grade level (i.e., elementary cohort and middle school cohort) and subgroup.

Houston Independent School District: Results for Spring 2014 State Assessments

Houston Independent School District (HISD)
Spring 2014 State of Texas Assessment of Academic Readiness (STAAR)
Key Findings for Phase 1

For all students combined (the "All" group) and the specified subgroups in the Houston region, the following outcomes favoring Phase 1 students were found on the Spring 2014 STAAR reading/mathematics/science.

IEP

- <u>Elementary Cohort in science</u>: Phase 1 had a substantively higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.37). It should be noted that the samples sizes for both Phase 1 (n =21) and Phase 2 (n =18) were small.
- <u>Middle School Cohort in mathematics</u>: Phase 1 had a substantively higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.82). However, the sample sizes for both Phase 1 (n = 3) and Phase 2 (n = 4) were very small.

ELL

- <u>Middle School Cohort in mathematics</u>: Phase 1 had a substantively higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.46). It should be noted that the sample size for Phase 1 (n = 17) was small.

Houston Independent School District (HISD): Spring 2014 State of Texas Assessments of Academic Readiness (STAAR) Results

Houston Independent School District (HISD) State of Texas Assessments of Academic Readiness (STAAR) results from the Spring 2014 administrations are currently available and are reported below. It should be noted that as the PASS assessment is better aligned with and more sensitive to changes in program outcomes (i.e., inquiry-based science instruction and knowledge/application), results from the state assessments should be interpreted judiciously when being used to evaluate LASER program impacts.

Houston Independent School District (HISD): Elementary and Middle School Cohort State of Texas Assessments of Academic Readiness (STAAR) Spring 2014 Analyses

There were a total of 1,155 elementary cohort students in Phase 1 (n = 670) and Phase 2 (n = 485) schools and 245 middle school cohort students in Phase 1 (n = 132) and Phase 2 (n = 113) schools for the analysis of the HISD STAAR test in reading, and 1,054 elementary cohort students in Phase 1 (n = 600) and Phase 2 (n = 454) schools and 182 middle school cohort students in Phase 1 (n = 91) and Phase 2 (n = 91) schools for the analysis of the HISD STAAR test in mathematics, and 1,163 elementary cohort students in Phase 1 (n = 672) and Phase 2 (n = 491) schools and 243 middle school cohort students in Phase 1 (n = 131) and Phase 2 (n = 112) schools for the analysis of the HISD STAAR test in science. To be included in the analysis, a student had to meet two criteria: 1) a student had to have scores on the multiple choice sections of PASS in both Fall 2011 and Spring 2014, and 2) a student had to take the Spring 2014 STAAR reading, mathematics, or science assessment and the selected baseline achievement assessment. With respect to the students included in the analysis, hierarchical or “block entry” multiple regressions were conducted to determine whether groups of students within cohort grade levels differed by Phase in their performance on 2013-2014 STAAR reading, mathematics and science scaled scores. In addition to these regressions, a second set of analyses (ANCOVA) intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (g) were also conducted on the outcomes for all students by Phase within cohort grade level, as well as for subgroups of these same students, categorized by their IEP (Special Education) status, ELL (English Language Learner) status, Economically Disadvantaged (FRL) status, and Gender. As the analyses were all exploratory in nature, no corrections were made for multiple comparisons.

In the selection of the baseline achievement test, four major factors were considered: (1) the number of students available for analysis; (2) the correlation between the baseline and current test scores; (3) whether or not the ANCOVA assumption of homogeneity of variance was met; and (4) independent t-test results (i.e., whether or not there was a non-significant difference in the baseline achievement between Phase 1 and Phase 2 students overall and by subgroups).

It should be noted that because students in the elementary cohort do not have a baseline-year (i.e., pre-program) STAAR test score available in either reading or mathematics or science, the Fall 2011 PASS scaled score and the 2010-2011 mathematics Stanford NCE score were used as the prior-achievement measures for the reading, mathematics, and science analyses respectively. The correlation between the Fall 2011 PASS scaled score and the 2013-2014 reading STAAR scaled score was moderately strong and statistically significant (r = 0.59, p < 0.001). The correlation between the 2010-2011 mathematics Stanford NCE score and the 2013-2014 mathematics STAAR scaled score was low, and also statistically significant ($r = 0.47$, $p < 0.001$). The correlation between the Fall 2011 PASS scaled score and the 2013-2014 science STAAR scaled score was moderately strong and statistically significant ($r = 0.58$, $p < 0.001$).

As the state assessment in Texas changed from the Texas Assessment of Knowledge and Skills (TAKS) to the STAAR between the 2010-2011 and 2012-2013 school years, students in the middle school cohort did not have a baseline-year STAAR test score available in either reading, mathematics or science. Therefore, the 2010-2011 mathematics TAKS scaled score was used as the prior-achievement measures for the reading, mathematics and science analyses. Correlation between the 2010-2011 mathematics TAKS scaled score and the 2013-2014 reading STAAR scaled score was moderate and statistically significant ($r = 0.53$, $p < 0.001$). Correlation between the 2010-2011 mathematics TAKS scaled score and the 2013-2014 mathematics STAAR scaled score was also moderate and statistically significant ($r = 0.60$, $p < 0.001$). Correlation between the 2010-2011 mathematics TAKS scaled score and the 2013-2014 science STAAR scaled score was also moderate and statistically significant ($r = 0.63$, $p < 0.001$).

To determine baseline achievement score equivalence between Phase 1 and Phase 2 students included in the present analysis, a series of independent t-tests was conducted for all elementary and middle school cohort students in the aggregate as well as for subgroups of these students by their Special Education (IEP) status, English language learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender. In addition, an effect size was also calculated as a measure of baseline equivalence.

As an indicator of the impact or “practical significance” of the treatment, the “effect size” (calculated as Hedges’s *g*) is a descriptive statistic that indicates the magnitude of the difference (in standard deviation units) between two measures. For example, a positive effect size would indicate a higher (i.e., better) Phase 1 mean, while a negative effect size would indicate a higher (i.e., better) Phase 2 mean. Based on guidelines from the What Works Clearinghouse (WWC), part of the research arm of the U.S. Department of Education, an effect size of +/- 0.25 is considered to be “substantively important”. As the analyses were all exploratory in nature, no corrections were made for multiple comparisons.

As shown in Table 1, for the elementary cohort in reading, neither statistically significant nor substantively important differences by phase in the baseline achievement levels were found for students in the aggregate (the “All” group). However, statistically significant differences were found for students in three subgroups (Not ELL: $t(557) = 3.31$, $p = 0.001$, $g = 0.29$, $PR = 61$; ELL: $t(594) = -2.11$, $p = 0.035$, $g = -0.18$, $PR = 43$; Female: $t(583) = 2.10$, $p = 0.036$, $g = 0.18$, $PR = 57$), with the difference between Phase 1 and Phase 2 in the Not ELL subgroup being substantively important according to the WWC guideline. Specifically, Phase 1 students were favored in the Not ELL subgroup. For the elementary cohort in mathematics (see Table 2), again, neither statistically significant nor substantively important differences by phase in the baseline achievement levels were found for students in the aggregate. Statistically significant differences were found for students in subgroups ELL ($t(556) = 2.68$, $p = 0.008$, $g = 0.23$, $PR = 59$) and FRL ($t(896) = 1.98$, $p = 0.048$, $g = 0.13$, $PR = 55$). However, the effect sizes associated with the Phase 1 and Phase 2 differences in the ELL and FRL subgroups did not meet the WWC threshold for substantive importance. For the elementary cohort in science (see Table 3), just like in reading and mathematics, neither statistically significant nor substantively important differences by phase in the baseline achievement levels were found for students in the aggregate. Like the reading test, statistically significant differences were found for students in three subgroups (Not ELL: $t(559) = 3.20$, $p = 0.001$, $g = 0.28$, $PR = 61$; ELL: $t(600) = -2.00$, $p = 0.046$, $g = -0.16$, $PR = 44$; Female: $t(582) = 2.07$, $p = 0.039$, $g = 0.17$, $PR = 57$), and the difference between Phase 1 and Phase 2 in the Not ELL subgroup was substantively important, with the phase 1 students being favored.

Table 1. STAAR Reading, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – Fall 2011 PASS-B Scaled Scores (*N* = 1,155)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	670	301.7	102.50	485	293.3	92.39	1.45	0.09	54
Not IEP	651	303.1	102.10	472	294.1	91.20	1.53	0.10	54
IEP	19	256	109.70	13	264.2	129.90	-0.19	-0.06	48
Not ELL	343	331.7	103.00	216	302.5	99.50	3.31**	0.29	61
ELL	327	270.3	92.22	269	285.8	85.73	-2.11*	-0.18	43
Not FRL	114	370.2	115.70	54	353.6	111.60	0.88	0.15	56
FRL	556	287.7	93.69	431	285.7	86.94	0.34	0.02	51
Male	345	296.9	106.50	225	296.8	92.10	0.01	0.00	50
Female	325	306.9	98.03	260	290.2	92.71	2.10*	0.18	57

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Table 2. STAAR Mathematics, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2010-2011 Stanford Mathematics NCE (*N* = 1,054)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	600	67.5	21.12	454	65.4	21.39	1.60	0.10	54
Not IEP	580	68.0	20.93	440	65.8	21.20	1.66	0.10	54
IEP	20	50.9	20.36	14	50.5	22.60	0.05	0.02	51
Not ELL	303	58.6	16.91	193	56.8	19.17	1.14	0.11	54
ELL	297	76.5	21.19	261	71.7	20.73	2.68**	0.23	59
Not FRL	105	64.8	17.40	51	67.1	18.53	-0.76	-0.13	45
FRL	495	68.0	21.80	403	65.1	21.73	1.98*	0.13	55
Male	314	67.4	21.41	217	66.3	21.93	0.58	0.05	52
Female	286	67.5	20.83	237	64.5	20.89	1.65	0.14	56

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$

Table 3. STAAR Science, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) - Fall 2011 PASS-B Scaled Scores (N = 1,163)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	672	301.0	102.50	491	292.4	92.31	1.47	0.09	53
Not IEP	651	302.7	102.10	473	293.6	91.33	1.53	0.09	54
IEP	21	250.4	106.00	18	260.9	113.70	-0.30	-0.09	46
Not ELL	344	330.6	103.20	217	302.4	99.41	3.20**	0.28	61
ELL	328	270.0	92.26	274	284.6	85.64	-2.00*	-0.16	44
Not FRL	113	369.8	116.20	55	351.3	111.90	0.98	0.16	56
FRL	559	287.1	93.72	436	285.0	86.89	0.36	0.02	51
Male	349	296.4	106.20	230	295.8	91.79	0.07	0.01	50
Female	323	306.1	98.37	261	289.5	92.84	2.07*	0.17	57

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$

For the middle school cohort in reading (see Table 4), no statistically significant differences by phase in the baseline achievement levels were found for students in either the aggregate or any subgroups. Although no statistically significant differences were found, the effect sizes associated with the differences in the subgroups IEP (t (4) = 0.58, p = 0.594, g = 0.38, PR = 65), ELL (t (59) = 1.89, p = 0.064, g = 0.48, PR = 68), and Not FRL (t (16) = -0.39, p = 0.700, g = -0.28, PR = 39) met the WWC threshold for substantive importance, favoring Phase 1 students in the IEP and ELL subgroups and Phase 2 students in the Not FRL subgroup. For the middle school cohort in mathematics (see Table 5), no statistically significant differences were found for students either in the aggregate or by subgroup, and the effect sizes did not meet the WWC threshold for substantive importance for any comparisons. For the middle school cohort in science (see Table 6), like in reading, while no statistically significant differences by phase in the baseline achievement levels were found for students in either the aggregate or any subgroups, the differences in the subgroups IEP (t (4) = 0.58, p = 0.594, g = 0.38, PR = 65), ELL (t (58) = 1.70, p = 0.094, g = 0.44, PR = 67), and Not FRL (t (16) = -0.39, p = 0.700, g = -0.28, PR = 39) were substantively important, favoring Phase 1 students in the IEP and ELL subgroups and Phase 2 students in the Not FRL subgroup.

Therefore, the outcomes should be interpreted cautiously in light of the substantively important differences in baseline achievement between Phase 1 and Phase 2 students for the following subgroups: Not ELL elementary cohort students in reading and science (favoring Phase 1), and the middle school cohort students in the IEP and ELL subgroups (both favoring Phase 1) and the Not FRL subgroup (favoring Phase 2) in reading and science. Note that the sample sizes for the Phase 1 (n = 3) and Phase 2 (n = 3) Not FRL subgroups in the middle school cohort were both very small, and may not be representative of those subgroups.

Table 4. STAAR Reading, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2010-2011 TAKS Mathematics Scaled Scores (*N* = 245)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	132	713.8	103.00	113	702.2	95.64	0.91	0.12	55
Not IEP	129	714.8	103.00	110	704.2	95.78	0.82	0.11	54
IEP	3	671.0	116.00	3	627.3	60.25	0.58	0.38	65
Not ELL	105	711.5	105.50	79	713.6	92.39	-0.14	-0.02	49
ELL	27	723.0	94.07	34	675.8	99.21	1.89	0.48	68
Not FRL	16	737.1	88.36	2	765.5	180.30	-0.39	-0.28	39
FRL	116	710.6	104.80	111	701.1	94.57	0.72	0.09	54
Male	65	718.0	96.02	47	706.3	96.33	0.64	0.12	55
Female	67	709.7	109.90	66	699.3	95.78	0.59	0.10	54

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Table 5. STAAR Mathematics, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2010-2011 TAKS Mathematics Scaled Scores (*N* = 182)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	91	672.6	84.44	91	686.0	95.02	-1.01	-0.03	49
Not IEP	88	672.7	84.07	87	689.8	94.60	-1.27	-0.04	49
IEP	3	671.0	116.00	4	603.3	68.85	0.98	0.13	55
Not ELL	74	672.6	87.39	61	695.2	93.37	-1.45	-0.04	48
ELL	17	672.8	72.55	30	667.5	97.21	0.20	0.01	51
Not FRL	7	687.3	56.86	2	765.5	180.30	-1.13	-0.11	46
FRL	84	671.4	86.48	89	684.2	93.37	-0.94	-0.03	49
Male	43	679.1	80.47	41	682.1	90.38	-0.16	-0.01	50
Female	48	666.9	88.29	50	689.3	99.45	-1.18	-0.05	48

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Table 6. STAAR Science, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2010-2011 TAKS Mathematics Scaled Scores (*N* = 243)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	131	714.9	102.70	112	703.9	94.29	0.86	0.11	54
Not IEP	128	715.9	102.70	109	706.0	94.35	0.76	0.10	54
IEP	3	671.0	116.00	3	627.3	60.25	0.58	0.38	65
Not ELL	104	712.8	105.20	79	713.6	92.39	-0.05	-0.01	50
ELL	27	723.0	94.07	33	680.9	96.20	1.70	0.44	67
Not FRL	16	737.1	88.36	2	765.5	180.30	-0.39	-0.28	39
FRL	115	711.8	104.50	110	702.8	93.19	0.68	0.09	54
Male	64	720.2	95.13	47	706.3	96.33	0.76	0.14	56
Female	67	709.7	109.90	65	702.2	93.50	0.42	0.07	53

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

STAAR Spring 2014 Results: Elementary Cohort Reading

For the 1,155 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2011 PASS-Basic scaled scores (Block 3) explained 38% of the total variance (R^2) in students' 2013-2014 reading STAAR scaled scores (see Table 7). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 reading scaled scores (β = -0.01, t = -0.59, p = 0.558).

The overall ANCOVA analysis (see table 8) revealed that there was a neither statistically significant nor substantively important difference in students' 2013-2014 reading STAAR scaled scores between Phase 1 and Phase 2 elementary cohort students overall. Consistent with the overall outcome, all subgroup ANCOVA analyses revealed neither statistically significant nor substantively important differences.

Table 7. STAAR Reading, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 Scaled Scores (N = 1,155)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1150) = 40.53$, $p < 0.001$, $R^2 = 0.124$ F Change $(4,1150) = 40.53$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-84.70	20.16	-0.12	-4.20	< 0.001***
ELL (0 = No, 1 = ELL)	-42.14	7.01	-0.18	-6.01	< 0.001***
FRL (0 = No, 1 = FRL)	-79.39	9.93	-0.23	-7.99	< 0.001***
Gender (0 = M, 1= F)	11.38	6.63	0.05	1.72	0.086
Block 2: Demographics + Fall 2011 PASS Scaled Score Model Fit: $F(5, 1149) = 141.17$, $p < 0.001$, $R^2 = 0.381$ F Change $(1,1149) = 476.69$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-48.74	17.04	-0.07	-2.86	0.004**
ELL (0 = No, 1 = ELL)	-23.72	5.95	-0.10	-3.98	< 0.001***
FRL (0 = No, 1 = FRL)	-35.81	8.59	-0.11	-4.17	< 0.001***
Gender (0 = M, 1= F)	10.24	5.57	0.04	1.84	0.066
Fall 2011 PASS Scaled Score	0.65	0.03	0.54	21.83	< 0.001***
Block 3: Demographics + Fall 2011 PASS Scaled Score + Phase Model Fit: $F(6, 1148) = 117.63$, $p < 0.001$, $R^2 = 0.381$ F Change $(1,1148) = 0.34$, $p = 0.558$					
IEP (0 = No, 1 = IEP)	-48.77	17.04	-0.07	-2.86	0.004**
ELL (0 = No, 1 = ELL)	-23.87	5.96	-0.10	-4.00	< 0.001***
FRL (0 = No, 1 = FRL)	-36.09	8.61	-0.11	-4.19	< 0.001***
Gender (0 = M, 1= F)	10.07	5.58	0.04	1.80	0.071
Fall 2011 PASS Scaled Score	0.65	0.03	0.54	21.83	< 0.001***
Phase (0 = P2, 1 = P1)	-3.31	5.65	-0.01	-0.59	0.558

** $p < 0.01$; *** $p < 0.001$

Table 8. STAAR Reading, HISD, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 1,155)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	670	1534.0	123.20	1530.4	485	1528.7	114.00	1533.7	0.34	0.558	-0.03	49
Not IEP	651	1535.7	123.42	1531.9	472	1530.7	113.65	1535.9	0.50	0.482	-0.03	49
IEP	19	1474.8	101.08	1474.6	13	1455.9	106.34	1456.2	0.27	0.605	0.17	57
Not ELL	343	1565.8	119.32	1557.2	216	1557.7	114.66	1571.3	3.19	0.075	-0.12	45
ELL	327	1500.6	118.44	1505.4	269	1505.3	108.13	1499.5	0.52	0.471	0.05	52
Not FRL	114	1617.24	118.11	1613.5	54	1604.3	125.19	1612.1	0.01	0.916	0.01	50
FRL	556	1516.9	117.21	1516.2	431	1519.2	109.03	1520.1	0.42	0.519	-0.04	49
Male	345	1524.1	127.62	1522.3	225	1526.5	117.80	1529.2	0.70	0.404	-0.06	48
Female	325	1544.5	117.61	1538.2	260	1530.6	110.79	1538.4	0.00	0.973	0.00	50

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

STAAR Spring 2014 Results: Elementary Cohort Mathematics

For the 1,054 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 mathematics Stanford NCE scores (Block 3) explained 31% of the total variance (R^2) in students' 2013-2014 mathematics STAAR scaled scores (see Table 9). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 mathematics scaled scores (β = -0.03, t = -1.03, p = 0.302).

The overall ANCOVA analysis (see Table 10) revealed that there was no statistically significant difference between Phase 1 and Phase 2 elementary cohort students' 2013-2014 mathematics STAAR scaled scores overall, and the effect size (g = -0.05) favoring Phase 2 students was not substantively important according to WWC guidelines.

The ANCOVA analyses for the subgroup comparisons revealed that Phase 1 students statistically significantly outperformed their Phase 2 counterparts in the Not ELL and Not FRL subgroups, whereas Phase 2 students statistically significantly outperformed their Phase 1 counterparts in the ELL and FRL subgroups. In addition, the effect size associated with the Not FRL (g = 0.35) subgroup comparison was substantively important, with the average Phase 1 Not FRL student scoring at the 64th percentile of the Not FRL control group (PR = 64). Given the statistical and substantive baseline equivalence between Phase 1 and Phase 2 students within the Not FRL subgroup, it appears that Phase 1 Not FRL elementary cohort students achieved advantages on the 2014 STAAR mathematics compared to their Phase 2 counterparts. No other subgroup comparisons reached the WWC threshold for substantive importance, ranging from -0.21 (ELL) to 0.19 (IEP).

Table 9. STAAR Mathematics, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 Scaled Scores (*N* = 1,054)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1049) = 17.17$, $p < 0.001$, $R^2 = 0.061$ F Change (4, 1049) = 17.17, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-100.31	23.61	-0.13	-4.25	< 0.001***
ELL (0 = No, 1 = ELL)	8.97	8.90	0.03	1.01	0.314
FRL (0 = No, 1 = FRL)	-86.76	12.52	-0.22	-6.93	< 0.001***
Gender (0 = M, 1= F)	18.46	8.36	0.07	2.21	0.027*
Block 2: Demographics + 2010-2011 Stanford mathematics NCE Score Model Fit: $F(5, 1048) = 93.49$, $p < 0.001$, $R^2 = 0.308$ F Change (1,1048) = 374.31, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-43.32	20.49	-0.06	-2.11	0.035*
ELL (0 = No, 1 = ELL)	-56.48	8.36	-0.20	-6.76	< 0.001***
FRL (0 = No, 1 = FRL)	-57.06	10.86	-0.15	-5.25	< 0.001***
Gender (0 = M, 1= F)	17.25	7.18	0.06	2.40	0.016*
2010-2011 Stanford mathematics NCE Score	3.59	0.19	0.55	19.35	< 0.001***
Block 3: Demographics + 2010-2011 Stanford mathematics NCE Score + Phase Model Fit: $F(6, 1047) = 78.09$, $p < 0.001$, $R^2 = 0.309$ F Change (1,1047) = 1.07, $p = 0.302$					
IEP (0 = No, 1 = IEP)	-43.12	20.49	-0.05	-2.10	0.036*
ELL (0 = No, 1 = ELL)	-57.21	8.39	-0.21	-6.82	< 0.001***
FRL (0 = No, 1 = FRL)	-57.62	10.88	-0.15	-5.30	< 0.001***
Gender (0 = M, 1= F)	16.90	7.19	0.06	2.35	0.019*
2010-2011 Stanford mathematics NCE Score	3.60	0.19	0.55	19.37	< 0.001***
Phase (0 = P2, 1 = P1)	-7.52	7.27	-0.03	-1.03	0.302

* $p < 0.05$; *** $p < 0.001$

Table 10. STAAR Mathematics, HISD, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (*N* = 1,054)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	600	1630.3	140.95	1623.9	454	1622.9	136.29	1631.4	1.07	0.302	-0.05	48
Not IEP	580	1633.0	140.95	1626.6	440	1626.2	134.49	1634.7	1.20	0.275	-0.06	48
IEP	20	1553.0	120.14	1550.2	14	1519.8	156.89	1523.7	0.42	0.525	0.19	58
Not ELL	303	1646.3	134.64	1641.9	193	1614.0	130.24	1620.9	4.78	0.029*	0.16	56
ELL	297	1614.0	145.53	1606.9	261	1629.5	140.48	1637.6	8.36	0.004**	-0.21	42
Not FRL	105	1706.1	136.72	1708.4	51	1666.8	120.43	1662.0	5.79	0.017*	0.35	64
FRL	495	1614.2	136.66	1608.6	403	1617.4	137.30	1624.3	4.04	0.045*	-0.11	45
Male	314	1621.5	133.15	1616.0	217	1614.3	131.18	1622.3	0.41	0.524	-0.05	48
Female	286	1640.0	148.68	1631.4	237	1630.8	140.62	1641.1	0.82	0.366	-0.07	47

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60^{th} percentile of the control group.
* $p < 0.05$; ** $p < 0.01$

STAAR Spring 2014 Results: Elementary Cohort Science

For the 1,163 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2011 PASS-Basic scaled scores (Block 3) explained 36% of the total variance (R^2) in students' 2013-2014 science STAAR scaled scores (see Table 11). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 science scaled scores ($\beta = 0.04$, $t = 1.75$, $p = 0.080$).

The overall ANCOVA analysis (see Table 12) revealed that there was no statistically significant difference between Phase 1 and Phase 2 elementary cohort students' 2013-2014 science STAAR scaled scores overall, and the effect size ($g = 0.08$) favoring Phase 1 students was not substantively important according to WWC guidelines.

The ANCOVA analyses for the subgroup comparisons revealed that Phase 1 female students statistically significantly outperformed Phase 2 female students. However, the effect size associated with this subgroup comparison ($g = 0.17$) was not substantively important, with the average Phase 1 female students scoring at the 57th percentile of the Phase 2 female students ($PR = 57$). Although not statistically significant, the effect size associated with the IEP subgroup ($g = 0.37$) comparison was substantively important, with the average Phase 1 IEP students scoring at the 65th percentile of the Phase 2 IEP students ($PR = 65$). However, the sample sizes for both Phase 1 ($n = 21$) and Phase 2 ($n = 18$) were small. Given the statistical and substantive baseline equivalence between Phase 1 and Phase 2 students within the IEP subgroup, it appears that Phase 1 IEP elementary cohort students achieved advantages on the 2014 STAAR science compared to their Phase 2 counterparts. All other subgroup comparisons were neither statistically significant nor substantively important, with effect size ranging from 0.02 (Male) to 0.17 (Female).

Table 11. STAAR Science, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 Scaled Scores (N = 1,163)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1158) = 30.92$, $p < 0.001$, $R^2 = 0.096$ F Change (4, 1158) = 30.92, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-301.51	68.48	-0.12	-4.40	< 0.001***
ELL (0 = No, 1 = ELL)	-60.33	26.09	-0.07	-2.31	0.021*
FRL (0 = No, 1 = FRL)	-317.34	37.10	-0.25	-8.55	< 0.001***
Gender (0 = M, 1= F)	-69.41	24.68	-0.08	-2.81	0.005**
Block 2: Demographics + Fall 2011 PASS Scaled Score Model Fit: $F(5, 1157) = 127.39$, $p < 0.001$, $R^2 = 0.355$ F Change (1,1157) = 463.85, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-166.87	58.22	-0.07	-2.87	0.004**
ELL (0 = No, 1 = ELL)	7.97	22.28	0.01	0.36	0.721
FRL (0 = No, 1 = FRL)	-157.55	32.22	-0.13	-4.89	< 0.001***
Gender (0 = M, 1= F)	-72.70	20.86	-0.08	-3.48	0.001**
Fall 2011 PASS Scaled Score	2.40	0.11	0.54	21.54	< 0.001***
Block 3: Demographics + Fall 2011 PASS Scaled Score + Phase Model Fit: $F(6, 1156) = 106.86$, $p < 0.001$, $R^2 = 0.357$ F Change (1,1156) = 3.08, $p = 0.080$					
IEP (0 = No, 1 = IEP)	-164.56	58.18	-0.07	-2.83	0.005**
ELL (0 = No, 1 = ELL)	9.84	22.29	0.01	0.44	0.659
FRL (0 = No, 1 = FRL)	-154.54	32.24	-0.12	-4.79	< 0.001***
Gender (0 = M, 1= F)	-70.76	20.87	-0.08	-3.39	0.001**
Fall 2011 PASS Scaled Score	2.40	0.11	0.54	21.53	< 0.001***
Phase (0 = P2, 1 = P1)	37.04	21.12	0.04	1.75	0.080

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 12. STAAR Science, HISD, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 1,163)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	672	3812.8	460.22	3798.8	491	3742.7	407.00	3761.8	3.08	0.080	0.08	53
Not IEP	651	3820.0	461.42	3805.8	473	3751.6	399.90	3771.1	2.59	0.108	0.08	53
IEP	21	3588.4	362.91	3629.9	18	3509.3	523.92	3460.8	2.32	0.137	0.37	65
Not ELL	344	3885.6	455.91	3852.2	217	3787.8	427.78	3840.8	0.15	0.700	0.03	51
ELL	328	3736.4	452.93	3749.4	274	3707.0	386.83	3691.3	3.68	0.055	0.14	55
Not FRL	113	4107.8	502.21	4090.7	55	3999.0	506.80	4034.0	0.95	0.332	0.11	54
FRL	559	3753.2	427.65	3749.2	436	3710.4	381.26	3715.5	2.22	0.137	0.08	53
Male	349	3822.8	467.23	3820.0	230	3808.8	414.86	3813.1	0.05	0.823	0.02	51
Female	323	3801.9	452.99	3782.0	261	3684.4	391.56	3709.1	6.14	0.014*	0.17	57

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (g). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$

STAAR Spring 2014 Results: Middle School Cohort Reading

For the 245 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 mathematics TAKS scaled scores (Block 3) explained 31% of the total variance (R^2) in students' 2013-2014 reading STAAR scaled scores (see Table 13). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 reading scaled scores ($\beta = 0.04$, $t = 0.72$, $p = 0.472$).

The overall ANCOVA analysis (see Table 14) revealed that there was no statistically significant difference between Phase 1 and Phase 2 elementary cohort students' 2013-2014 science STAAR scaled scores overall, and the effect size ($g = 0.08$) favoring Phase 1 students was not substantively important according to WWC guidelines.

The ANCOVA analyses for the subgroup comparisons revealed that Phase 1 students statistically significantly outperformed their Phase 2 counterparts in the ELL ($g = 0.59$) subgroup. The effect size associated with the difference for the ELL subgroup was also substantively important according to WWC guidelines, and favored Phase 1 students, with the average Phase 1 student scoring at the 72nd percentile of the control group ($PR = 72$). However, Phase 1 ELL students had a substantively important advantage on the pretest ($g = 0.48$). Therefore, the large effect sizes for the posttest (i.e., spring 2014 STAAR reading) could be a function of the large advantage they had at the pretest, and appears to indicate that Phase 1 students maintained their pretest advantage by spring 2014. In addition, although Not FRL Phase 1 students did not have a statistically significantly higher adjusted mean compared to their Phase 2 counterparts, the effect size for the Not FRL subgroup ($g = 0.92$) was substantively important, with the average Phase 1 student scoring at the 82nd percentile of the control group ($PR = 82$). It should be noted that for the Not FRL subgroup, Phase 2 students had substantively higher baseline scores ($g = -0.28$). Therefore, it appears that Phase 1 Not FRL students were able to not only greatly reduce, but even to reverse the achievement gap present at the baseline. However, we should also note that the small sample sizes for the Not FRL subgroup, particularly for Phase 2 students ($n = 2$) would indicate that this outcome would not be representative of this subgroups' performance. All other subgroup comparisons were neither statistically significant nor substantively important, with the effect size ranging from -0.11 (Not ELL) to 0.17 (Female).

Table 13. STAAR Reading, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 Scaled Scores (*N* = 245)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 240) = 3.33$, $p = 0.011$, $R^2 = 0.053$ F Change $(4, 240) = 3.33$, $p = 0.011$					
IEP (0 = No, 1 = IEP)	-132.45	45.59	-0.18	-2.91	0.004**
ELL (0 = No, 1 = ELL)	-14.37	16.29	-0.06	-0.88	0.379
FRL (0 = No, 1 = FRL)	-44.37	26.84	-0.10	-1.65	0.100
Gender (0 = M, 1= F)	17.76	14.11	0.08	1.26	0.209
Block 2: Demographics + 2010-2011 TAKS mathematics Scaled Scores Model Fit: $F(5, 239) = 21.59$, $p < 0.001$, $R^2 = 0.311$ F Change $(1, 239) = 89.73$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-91.37	39.20	-0.13	-2.33	0.021*
ELL (0 = No, 1 = ELL)	-3.65	13.96	-0.01	-0.26	0.794
FRL (0 = No, 1 = FRL)	-23.82	23.03	-0.06	-1.03	0.302
Gender (0 = M, 1= F)	24.01	12.08	0.11	1.99	0.048*
2010-2011 TAKS mathematics Scaled Scores	0.58	0.06	0.52	9.47	< 0.001***
Block 3: Demographics + 2010-2011 TAKS mathematics Scaled Scores + Phase Model Fit: $F(6, 238) = 18.05$, $p < 0.001$, $R^2 = 0.313$ F Change $(1, 238) = 0.52$, $p = 0.472$					
IEP (0 = No, 1 = IEP)	-90.31	39.26	-0.13	-2.30	0.022*
ELL (0 = No, 1 = ELL)	-2.46	14.07	-0.01	-0.17	0.862
FRL (0 = No, 1 = FRL)	-20.53	23.50	-0.05	-0.87	0.383
Gender (0 = M, 1= F)	24.75	12.13	0.11	2.04	0.042*
2010-2011 TAKS mathematics Scaled Scores	0.58	0.06	0.51	9.44	< 0.001***
Phase (0 = P2, 1 = P1)	8.92	12.36	0.04	0.72	0.472

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 14. STAAR Reading, HISD, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (*N* = 245)

Group	Treatment (Phase 1)				Control (Phase 2)							
	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	132	1681.3	121.75	1677.8	113	1664.9	97.87	1668.9	0.52	0.472	0.08	53
Not IEP	129	1685.0	120.70	1681.8	110	1667.4	97.53	1671.1	0.73	0.393	0.10	54
IEP	3	1523.0	0.00	1549.3	3	1574.0	72.58	1547.7	0.00	0.968	0.03	51
Not ELL	105	1671.8	115.52	1671.7	79	1683.3	99.59	1683.4	0.71	0.399	-0.11	46
ELL	27	1718.2	139.73	1701.2	34	1622.1	79.81	1635.6	6.10	0.017*	0.59	72
Not FRL	16	1716.4	106.44	1722.1	2	1662.0	141.42	1616.3	3.23	0.097	0.92	82
FRL	116	1676.5	123.34	1673.9	111	1664.9	97.83	1667.5	0.25	0.617	0.06	52
Male	65	1666.7	117.46	1658.7	47	1656.4	114.96	1667.4	0.19	0.667	-0.07	47
Female	67	1695.5	125.01	1692.2	66	1670.9	84.03	1674.2	1.34	0.250	0.17	57

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$

STAAR Spring 2014 Results: Middle School Cohort Mathematics

For the 182 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 mathematics TAKS scaled scores (Block 3) explained 37% of the total variance (R^2) in students' 2013-2014 mathematics STAAR scaled scores (see Table 15). The addition of the student's Phase to the model (Block 3) did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 mathematics scaled scores (β= 0.04, t = 0.64, p = 0.522).

The overall ANCOVA analysis (see Table 16) revealed that there was no statistically significant difference in students' 2013-2014 mathematics STAAR scaled scores between Phase 1 and Phase 2 middle school cohort students overall, and the effect size (g = 0.08) favoring Phase 1 students was not substantively important according to WWC guidelines.

Consistent with the overall outcome, all subgroup ANCOVA analyses revealed no statistically significant difference in students' 2013-2014 mathematics STAAR scaled scores. The effect sizes associated with the IEP (g = 0.82), ELL (g = 0.46), and Not FRL (g = 0.72) subgroup comparisons were substantively important, with the average IEP Phase 1 student scoring at the 79th percentile of the IEP control group (PR = 79), the average ELL Phase 1 student scoring at the 68th percentile of the ELL control group (PR = 68), and the average Not FRL Phase 1 student scoring at the 76th percentile of the Not FRL control group (PR = 76). Given the statistical and substantive baseline equivalence between Phase 1 and Phase 2 groups, ANCOVA analyses results seem to indicate that Phase 1 middle school cohort students in the IEP, ELL, and Not FRL subgroups achieved advantages on the 2014 STAAR mathematics relative to their respective Phase 2 counterparts. But it should be noted that the small sample sizes for the IEP (n = 3 for Phase 1 and n = 4 for Phase 2), ELL (n = 17 for Phase 1), and Not FRL (n = 7 for Phase 1 and n = 2 for Phase 2) subgroup would indicate that the outcomes would not be representative of both subgroups' performances. The other subgroup effect sizes ranged from -0.06 (Not ELL) to 0.16 (Male).

Table 15. STAAR Mathematics, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 Scaled Scores (N = 182)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 177) = 0.42$, $p = 0.796$, $R^2 = 0.009$ F Change $(4, 177) = 0.42$, $p = 0.796$					
IEP (0 = No, 1 = IEP)	-38.74	45.91	-0.06	-0.84	0.400
ELL (0 = No, 1 = ELL)	-5.01	20.12	-0.02	-0.25	0.804
FRL (0 = No, 1 = FRL)	-9.40	40.30	-0.02	-0.23	0.816
Gender (0 = M, 1= F)	14.66	17.59	0.06	0.83	0.406
Block 2: Demographics + 2010-2011 TAKS mathematics Scaled Scores Model Fit: $F(5, 176) = 20.89$, $p < 0.001$, $R^2 = 0.372$ F Change $(1, 176) = 101.83$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	6.64	36.92	0.01	0.18	0.858
ELL (0 = No, 1 = ELL)	7.48	16.11	0.03	0.46	0.643
FRL (0 = No, 1 = FRL)	13.25	32.24	0.02	0.41	0.682
Gender (0 = M, 1= F)	19.20	14.05	0.08	1.37	0.173
2010-2011 TAKS mathematics Scaled Scores	0.79	0.08	0.61	10.09	< 0.001***
Block 3: Demographics + 2010-2011 TAKS mathematics Scaled Scores + Phase Model Fit: $F(6, 175) = 17.42$, $p < 0.001$, $R^2 = 0.374$ F Change $(1, 175) = 0.412$, $p = 0.522$					
IEP (0 = No, 1 = IEP)	8.44	37.08	0.01	0.23	0.820
ELL (0 = No, 1 = ELL)	9.29	16.38	0.04	0.57	0.571
FRL (0 = No, 1 = FRL)	15.94	32.57	0.03	0.49	0.625
Gender (0 = M, 1= F)	19.63	14.09	0.08	1.39	0.165
2010-2011 TAKS mathematics Scaled Scores	0.79	0.08	0.61	10.09	< 0.001***
Phase (0 = P2, 1 = P1)	9.17	14.29	0.04	0.64	0.522

*** $p < 0.001$

Table 16. STAAR Mathematics, HISD, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 182)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	91	1666.7	117.52	1673.4	91	1670.9	115.89	1664.2	0.41	0.522	0.08	53
Not IEP	88	1665.9	117.91	1673.9	87	1674.8	116.71	1666.8	0.23	0.630	0.06	52
IEP	3	1690.3	125.54	1678.1	4	1584.8	45.74	1593.9	2.05	0.289	0.82	79
Not ELL	74	1657.9	120.78	1666.9	61	1684.7	103.24	1673.7	0.20	0.658	-0.06	48
ELL	17	1705.0	95.93	1702.2	30	1642.7	135.66	1644.3	2.83	0.100	0.46	68
Not FRL	7	1667.0	73.68	1689.4	2	1707.0	80.61	1628.7	0.59	0.499	0.72	76
FRL	84	1666.7	120.76	1673.0	89	1670.1	116.75	1664.1	0.37	0.544	0.08	53
Male	43	1666.7	100.67	1668.6	41	1652.7	124.12	1650.7	0.65	0.424	0.16	56
Female	48	1666.7	131.86	1676.5	50	1685.8	107.66	1676.4	0.00	0.994	0.00	50

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (g). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

STAAR Spring 2014 Results: Middle School Cohort Science

For the 243 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 mathematics TAKS scaled scores (Block 3) explained 43% of the total variance (R^2) in students' 2013-2014 science STAAR scaled scores (see Table 17). The addition of student's Phase to the model increased the percentage of variance accounted for from 41% to 43%, and phase was a statistically significant predictor of 2013-2014 science STAAR scaled scores (β = -0.14, t = -2.81, p = 0.005).

The overall ANCOVA analysis (see Table 18) revealed that there was a statistically significant and substantively important difference in students' 2013-2014 science STAAR scaled scores between Phase 1 and Phase 2 middle school cohort students in the aggregate ($F(1, 236) = 7.90$, p = 0.005, g = -0.29, *PR* = 39), with Phase 2 students being favored. The average Phase 1 students scored at the 39th percentile of the Phase 2 students.

The ANCOVA analyses for the subgroup comparisons revealed that Phase 2 students statistically significantly outperformed their Phase 1 counterparts in the Not IEP, Not ELL, FRL, and Female subgroups. Furthermore, the effect sizes associated with these comparisons were also substantively important:

- Not IEP (g = -0.27, *PR* = 40), with the average Phase 1 Not IEP students scoring at the 40th percentile of their Phase 2 counterparts;
- Not ELL (g = -0.38, *PR* = 35), with the average Phase 1 Not ELL students scoring at the 35th percentile of their Phase 2 counterparts;
- FRL (g= -0.29, *PR* = 38), with the average Phase 1 FRL students scoring at the 38th percentile of their Phase 2 counterparts;
- Female (g = -0.45, *PR* = 33), with the average Phase 1 female students scoring at the 33rd percentile of their Phase 2 counterparts.

In addition, although not statistically significant, the effect size associated with the IEP (g = -1.06) and Not FRL (g = -0.25) subgroup comparisons were substantively important, with the average Phase 1 IEP students scoring at the 14th percentile of the Phase 2 IEP students (*PR* = 14), and the average Phase 1 Not FRL students scoring at the 40th percentile of the Phase 2 Not FRL students (*PR* = 40). Again, it should be noted that the small sample sizes for the IEP (n = 3 for Phase 1 and n = 4 for Phase 2) and Not FRL (n = 16 for Phase 1 and n = 2 for Phase 2) subgroup would indicate that the outcomes would not be representative of both subgroups' performances.

Table 17. STAAR Science, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 Scaled Scores (N = 243)

Source	B	*S.E.B.*	β	t	p
Block 1: Demographics Model Fit: $F(4, 238) = 0.82$, $p = 0.511$, $R^2 = 0.014$ F Change $(4, 238) = 0.82$, $p = 0.511$					
IEP (0 = No, 1 = IEP)	-196.20	226.29	-0.06	-0.87	0.387
ELL (0 = No, 1 = ELL)	15.68	81.46	0.01	0.19	0.848
FRL (0 = No, 1 = FRL)	-185.79	133.21	-0.09	-1.39	0.164
Gender (0 = M, 1= F)	-57.00	70.39	-0.05	-0.81	0.419
Block 2: Demographics + 2010-2011 TAKS mathematics Scaled Scores Model Fit: $F(5, 237) = 32.65$, $p < 0.001$, $R^2 = 0.408$ F Change $(1, 237) = 157.77$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	54.37	176.83	0.02	0.31	0.759
ELL (0 = No, 1 = ELL)	72.78	63.41	0.06	1.15	0.252
FRL (0 = No, 1 = FRL)	-66.20	103.87	-0.03	-0.64	0.524
Gender (0 = M, 1= F)	-20.05	54.74	-0.02	-0.37	0.715
2010-2011 TAKS mathematics Scaled Scores	3.49	0.28	0.64	12.56	< 0.001***
Block 3: Demographics + 2010-2011 TAKS mathematics Scaled Scores + Phase Model Fit: $F(6, 236) = 29.32$, $p < 0.001$, $R^2 = 0.427$ F Change $(1,236) = 7.90$, $p = 0.005$					
IEP (0 = No, 1 = IEP)	36.59	174.42	0.01	0.21	0.834
ELL (0 = No, 1 = ELL)	53.53	62.88	0.04	0.85	0.395
FRL (0 = No, 1 = FRL)	-123.55	104.40	-0.06	-1.18	0.238
Gender (0 = M, 1= F)	-31.66	54.11	-0.03	-0.59	0.559
2010-2011 TAKS mathematics Scaled Scores	3.51	0.27	0.64	12.81	< 0.001***
Phase (0 = P2, 1 = P1)	-154.76	55.05	-0.14	-2.81	0.005**

** $p < 0.01$; *** $p < 0.001$

Table 18. STAAR Science, HISD, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 243)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	p	g	PR
All	131	3756.6	607.91	3734.2	112	3862.9	448.90	3889.0	7.90	0.005**	-0.29	39
Not IEP	128	3765.3	611.51	3743.2	109	3862.0	451.93	3887.9	6.75	0.010*	-0.27	40
IEP	3	3384.0	249.56	3421.1	3	3895.0	391.99	3857.9	1.07	0.490	-1.06	14
Not ELL	104	3719.1	580.52	3712.4	79	3904.5	444.17	3913.3	10.77	0.001**	-0.38	35
ELL	27	3901.0	696.74	3784.4	33	3763.2	451.23	3858.6	0.35	0.556	-0.13	45
Not FRL	16	3963.7	495.86	3958.9	2	4053.0	499.22	4091.6	0.16	0.693	-0.25	40
FRL	115	3727.8	618.23	3713.8	110	3859.4	449.72	3874.0	8.17	0.005**	-0.29	38
Male	64	3830.7	665.67	3800.5	47	3843.9	441.62	3885.0	0.75	0.387	-0.14	44
Female	67	3685.8	542.70	3668.3	65	3876.6	457.02	3894.6	13.06	< 0.001***	-0.45	33

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (g). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Houston Independent School District (HISD)
Spring 2014 Stanford Achievement Test
Key Findings for Phase 1

For all students combined (the "All" group) and the specified subgroups in the Houston region, the following outcomes favoring Phase 1 students were found on the Spring 2014 Stanford reading/mathematics/science.

All

- Middle School Cohort in mathematics: Phase 1 had a statistically significantly and substantively higher (i.e., educationally meaningful) adjusted mean score than Phase 2 in Spring 2014 (g = 0.30).
- Middle School Cohort in reading: Phase 1 had a statistically significant and nearly substantively higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.24).

Economically Disadvantaged (FRL)

- Middle School Cohort in mathematics: Phase 1 had a statistically significantly and substantively higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.29).

IEP

- Middle School Cohort in reading: While Phase 2 students had a substantively higher baseline score than Phase 1 students (g = -0.69), Phase 1 had a higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.08). It should be noted that the sample sizes for Phase 1 (n = 13) and Phase 2 (n = 21) were small.

Male

- Middle School Cohort in mathematics: Phase 1 had a statistically significantly and substantively higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.38).

Female

- Middle School Cohort in reading: Phase 1 had a statistically significantly and substantively higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.27).
- Middle School Cohort in mathematics: Phase 1 had a statistically significantly and substantively higher adjusted mean score than Phase 2 in Spring 2014 (g = 0.25).

Houston Independent School District (HISD):
Spring 2014 Stanford Achievement Test (Stanford) Results

Houston Independent School District (HISD) Stanford Achievement Test results from the Spring 2014 administrations are currently available and are reported below. It should be noted that as the PASS assessment is better aligned with and more sensitive to changes in program outcomes (i.e., inquiry-based science instruction and knowledge/application), results from the state assessments should be interpreted judiciously when being used to evaluate LASER program impacts.

Houston Independent School District (HISD): Elementary and Middle School Cohort Stanford Spring 2014 Analyses

There were a total of 1,189 elementary cohort students in Phase 1 (*n* = 688) and Phase 2 (*n* = 501) schools and 291 middle school cohort students in Phase 1 (*n* = 148) and Phase 2 (*n* = 143) schools for the analysis of the HISD Stanford test in reading, and 1,084 elementary cohort students in Phase 1 (*n* = 616) and Phase 2 (*n* = 468) schools and 244 middle school cohort students in Phase 1 (*n* = 131) and Phase 2 (*n* = 113) schools for the analysis of the HISD Stanford test in mathematics, and 1,189 elementary cohort students in Phase 1 (*n* = 688) and Phase 2 (*n* = 501) schools and 291 middle school cohort students in Phase 1 (*n* = 148) and Phase 2 (*n* = 143) schools for the analysis of the HISD Stanford test in science. To be included in the analysis, a student had to meet two criteria: 1) a student had to have scores on the multiple choice sections of PASS in both Fall 2011 and Spring 2014, and 2) a student had to take the Spring 2014 Stanford reading, mathematics, or science test and the selected baseline achievement assessment.

For both elementary and middle school cohort students in Phase 1 and Phase 2 schools, hierarchical or "block entry" multiple regressions were conducted by subject area to determine whether groups of students within grade levels differed by Phase in their performance on 2013-2014 HISD Stanford reading, mathematics and science normal curve equivalent (NCE)[1] scores. In addition to these regressions, a second set of analyses (ANCOVA) intended to generate pairs of adjusted NCE score means and to compute the treatment effect sizes (*g*) were also conducted on the outcomes for all students by Phase within grade level, as well as for subgroups of these same students, categorized by their IEP (Special Education) status, ELL (English Language Learner) status, Economically Disadvantaged (FRL) status, and Gender.

In the selection of the baseline achievement test, four major factors were considered: (1) the number of students available for analysis; (2) the correlation between the baseline and current test scores; (3) whether or not the ANCOVA assumption of homogeneity of variance was met; and (4) independent t-test results (i.e., whether or not there was a non-significant difference in the baseline achievement between Phase 1 and Phase 2 students overall and by subgroups). The pool of baseline achievement tests included Fall 2011 PASS scaled score, Spring 2012 PASS scaled score, 2010-2011 reading/mathematics/science Stanford NCE score, and the 2010-2011 reading/ mathematics/science TAKS scaled score.

[1] Although the Stanford and Aprenda tests are parallel in content, the Aprenda test is not a translation of the Stanford test (source: http://www.houstonisd.org/Page/59886).Therefore, NCE scores were used as the outcome measure to allow combining Stanford and Aprenda data for analysis, as NCE scores provide a common metric to put the two tests on a comparable scoring scale.

For the elementary cohort, the Fall 2011 PASS scaled score, 2010-2011 mathematics Stanford NCE score, and Fall 2011 PASS scaled score were selected as the baseline-achievement measures for the reading, mathematics, and science analyses respectively. Correlation with the Fall 2011 PASS scaled score was moderately strong and statistically significant for the 2013-2014 reading Stanford NCE score ($r = 0.64$, $p < 0.001$). Correlation with the 2010-2011 mathematics Stanford NCE score was moderate and statistically significant for the 2013-2014 mathematics Stanford NCE score ($r = 0.53$, $p < 0.001$). Correlation with the Fall 2011 PASS scaled score was moderate and statistically significant for the 2013-2014 science Stanford NCE score ($r = 0.54$, $p < 0.001$).

For the middle school cohort, Spring 2012 PASS scaled score was selected as the prior-achievement measure for the reading and science analyses, and the 2010-2011 mathematics Texas Assessment of Knowledge and Skills (TAKS) scaled score was used as the prior-achievement measures for the math analysis. Correlation with the Spring 2012 PASS scaled score was moderate and statistically significant for the 2013-2014 reading Stanford NCE score ($r = 0.67$, $p < 0.001$). Correlation with the 2010-2011 mathematics TAKS scaled score was high and statistically significant for the 2013-2014 mathematics Stanford NCE score ($r = 0.73$, $p < 0.001$). Correlation with the Spring 2012 PASS scaled score was high and also statistically significant for the 2013-2014 science Stanford NCE score ($r = 0.71$, $p < 0.001$).

To determine baseline achievement score equivalence between Phase 1 and Phase 2 students included in the present analysis, a series of independent t-tests was conducted for all elementary and middle school cohort students in the aggregate as well as for subgroups of these students by their Special Education (IEP) status, English language learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender. In addition, an effect size was also calculated as a measure of baseline equivalence. As an indicator of the impact or “practical significance” of the treatment, the “effect size” (calculated as Hedges’s *g*) is a descriptive statistic that indicates the magnitude of the difference (in standard deviation units) between two measures. For example, a positive effect size would indicate a higher (i.e., better) Phase 1 mean, while a negative effect size would indicate a higher (i.e., better) Phase 2 mean. Based on guidelines from the What Works Clearinghouse (WWC), part of the research arm of the U.S. Department of Education, an effect size of +/- 0.25 is considered to be “substantively important”. As the analyses were all exploratory in nature, no corrections were made for multiple comparisons.

For the elementary cohort in reading (see Table 19), Phase 1 students in the Not ELL subgroup had both statistically significantly and substantively higher baseline scores than their Phase 2 counterparts (t (573) $= 3.13$, $p = 0.002$, $g = 0.27$). No statistically significant or substantively important differences were found for students either in the aggregate (the “All” group) or for any other subgroups. For the elementary cohort in mathematics (see Table 20), statistically significant differences in baseline achievement between Phase 1 and Phase 2 were only found for students in the ELL subgroup (t (569) $= 2.90$, $p = 0.004$, $g = 0.24$), and the effect size associated with this difference was very close to the WWC threshold for substantive importance. Other comparisons were neither statistically significant nor substantively important. For the elementary cohort in science (see Table 21), as in reading, Phase 1 students in the Not ELL subgroup had both statistically significantly and substantively higher baseline scores than their Phase 2 counterparts (t (573) $= 3.13$, $p = 0.002$, $g = 0.27$). No other comparisons were found statistically significant or substantively important.

Table 19. Stanford Reading, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – Fall 2011 PASS Scaled Scores (*N* = 1,189)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	688	298.8	103.0	501	289.8	93.68	1.55	0.09	54
Not IEP	660	301.4	102.5	478	292.4	92.06	1.52	0.09	54
IEP	28	238.8	99.5	23	235.7	112.00	0.10	0.03	51
Not ELL	354	327.6	104.4	221	300.0	100.50	3.13**	0.27	61
ELL	334	268.4	92.4	280	281.8	87.24	-1.84	-0.15	44
Not FRL	114	370.2	115.7	55	351.3	111.90	1.01	0.16	57
FRL	574	284.6	94.2	446	282.2	88.41	0.42	0.03	51
Male	354	294.8	106.3	235	292.0	94.06	0.33	0.03	51
Female	334	303.1	99.4	266	287.8	93.48	1.92	0.16	56

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
** $p < 0.01$

Table 20. Stanford Mathematics, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2010-2011 Stanford Mathematics NCE Scores (*N* = 1,084)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	616	66.4	22.16	468	64.2	22.07	1.63	0.10	54
Not IEP	589	67.6	21.35	446	65.3	21.34	1.70	0.11	54
IEP	27	41.7	25.24	22	42.4	25.61	-0.10	-0.03	49
Not ELL	314	57.5	18.41	199	55.8	19.96	0.98	0.09	54
ELL	302	75.7	21.93	269	70.5	21.52	2.90**	0.24	60
Not FRL	107	64.0	18.25	53	65.6	19.67	-0.52	-0.09	47
FRL	509	67.0	22.88	415	64.1	22.37	1.94	0.13	55
Male	320	66.8	22.20	224	65.1	22.70	0.89	0.08	53
Female	296	66.1	22.15	244	63.5	21.49	1.38	0.12	55

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
** $p < 0.01$

Table 21. Stanford Science, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) - Fall 2011 PASS Scaled Scores (*N* = 1,189)

Group	Treatment (Phase 1)			Control (Phase 2)			*t*	*g*	PR
	n	*M*	*SD*	*n*	*M*	*SD*			
All	688	298.8	103.00	501	289.8	93.68	1.55	0.09	54
Not IEP	660	301.4	102.50	478	292.4	92.06	1.52	0.09	54
IEP	28	238.8	99.46	23	235.7	112.00	0.10	0.03	51
Not ELL	354	327.6	104.40	221	300.0	100.50	3.13**	0.27	61
ELL	334	268.4	92.41	280	281.8	87.24	-1.84	-0.15	44
Not FRL	114	370.2	115.70	55	351.3	111.90	1.01	0.16	57
FRL	574	284.6	94.18	446	282.2	88.41	0.42	0.03	51
Male	354	294.8	106.30	235	292.0	94.06	0.33	0.03	51
Female	334	303.1	99.44	266	287.8	93.48	1.92	0.16	56

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
** $p < 0.01$

For the middle school cohort in reading (see Table 22), statistically significant differences in baseline achievement between Phase 1 and Phase 2 were only found for students in the ELL subgroup (t (72) = 2.95, p = 0.004, g = 0.69), and the effect size associated with this difference was substantively important according to WWC guidelines, favoring Phase 1 students. In addition, although not statistically significant, the effect sizes associated with the difference between Phase 1 and Phase 2 in the IEP (t (32) = -1.99, p = 0.055, g = -0.69) and Not FRL (t (17) = 0.60, p = 0.558, g = 0.36) subgroups were substantively important, with Phase 2 students being favored in the IEP subgroup and Phase 1 students being favored in the Not FRL subgroup. For the middle school cohort in mathematics (see Table 23), no statistically significant differences were found for students either in the aggregate or for any subgroups, but the effect sizes associated with the difference for the three subgroups IEP, ELL and Not FRL met the WWC threshold for substantive importance, favoring Phase 1 students in the IEP (g = 0.63) and ELL (g = 0.44) subgroup and Phase 2 students in the Not FRL (g = 0.28) subgroup. For the middle school cohort in science (see Table 24), as in reading, only the difference in the ELL subgroup was statistically significant (t (72) = 2.95, p = 0.004, g = 0.69), but the effect sizes associated with differences for the three subgroups ELL (g = 0.69), IEP (g = -0.69), and Not FRL (g = 0.36) met the WWC threshold for substantive importance, favoring Phase 1 students in the ELL and Not FRL subgroups and Phase 2 students in the IEP subgroup.

Therefore, the outcomes should be interpreted in light of the substantively important difference in baseline achievement between Phase 1 and Phase 2 for the Not ELL elementary cohort students in reading and science, and the middle school cohort students in the IEP, ELL, and Not FRL subgroups in reading, mathematics, and science.

Table 22. Stanford Reading, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – Spring 2012 PASS Scaled Scores (*N* = 291)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	148	310.4	134.50	143	310.9	104.00	-0.04	0.00	50
Not IEP	135	324.7	129.80	122	325.6	99.06	-0.06	-0.01	50
IEP	13	162.2	87.31	21	225.7	92.31	-1.99	-0.69	25
Not ELL	118	295.8	126.10	99	322.5	108.40	-1.66	-0.22	41
ELL	30	368.0	152.50	44	285.0	89.29	2.95**	0.69	76
Not FRL	16	359.4	141.70	3	303.3	195.30	0.60	0.36	64
FRL	132	304.5	133.00	140	311.1	102.50	-0.46	-0.06	48
Male	71	322.7	136.90	64	299.7	117.10	1.04	0.18	57
Female	77	299.0	132.10	79	320.0	91.83	-1.16	-0.18	43

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
** $p < 0.01$

Table 23. Stanford Mathematics, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2010-2011 TAKS Mathematics Scaled Scores (*N* = 244)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	131	714.9	102.70	113	702.4	95.27	0.98	0.13	55
Not IEP	128	715.9	102.70	109	706.0	94.35	0.76	0.10	54
IEP	3	671.0	116.00	4	603.3	68.85	0.98	0.63	74
Not ELL	104	712.8	105.20	80	711.3	94.05	0.10	0.01	51
ELL	27	723.0	94.07	33	680.9	96.20	1.70	0.44	67
Not FRL	16	737.1	88.36	2	765.5	180.30	-0.39	-0.28	39
FRL	115	711.8	104.50	111	701.3	94.19	0.79	0.11	54
Male	64	720.2	95.13	48	702.7	98.60	0.95	0.18	57
Female	67	709.7	109.90	65	702.2	93.50	0.42	0.07	53

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Table 24. Stanford Science, HISD, Spring 2014: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – Spring 2012 PASS Scaled Scores (N = 291)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
All	148	310.4	134.50	143	310.9	104.00	-0.04	0.00	50
Not IEP	135	324.7	129.80	122	325.6	99.06	-0.06	-0.01	50
IEP	13	162.2	87.31	21	225.7	92.31	-1.99	-0.69	25
Not ELL	118	295.8	126.10	99	322.5	108.40	-1.66	-0.22	41
ELL	30	368.0	152.50	44	285.0	89.29	2.95**	0.69	76
Not FRL	16	359.4	141.70	3	303.3	195.30	0.60	0.36	64
FRL	132	304.5	133.00	140	311.1	102.50	-0.46	-0.06	48
Male	71	322.7	136.90	64	299.7	117.10	1.04	0.18	57
Female	77	299.0	132.10	79	320.0	91.83	-1.16	-0.18	43

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (g). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
** $p < 0.01$

Stanford Spring 2014 Results: Elementary Cohort Reading

For the 1,189 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2011 PASS-Basic scaled scores (Block 3) explained 45% of the total variance (R^2) in students' 2013-2014 reading Stanford NCE scores (see Table 25). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 reading NCE scores ($\beta = 0.00$, $t = -0.04$, $p = 0.969$).

The overall ANCOVA analysis (see Table 26) revealed that there was neither a statistically significant nor a substantively important difference in students' 2013-2014 reading Stanford NCE scores between Phase 1 and Phase 2 elementary cohort students overall. Consistent with the overall outcome, all subgroup ANCOVA analyses revealed neither statistically significant nor substantively important differences.

Table 25. Stanford Reading, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 NCE Scores (*N* = 1,189)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1184) = 59.41$, $p < 0.001$, $R^2 = 0.167$ F Change $(4,1184) = 59.41$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-192.77	25.57	-0.20	-7.54	< 0.001***
ELL (0 = No, 1 = ELL)	-57.52	10.98	-0.15	-5.24	< 0.001***
FRL (0 = No, 1 = FRL)	-158.35	15.68	-0.28	-10.10	< 0.001***
Gender (0 = M, 1= F)	25.59	10.39	0.07	2.46	0.014*
Block 2: Demographics + Fall 2011 PASS Scaled Score Model Fit: $F(5, 1183) = 196.55$, $p < 0.001$, $R^2 = 0.454$ F Change $(1,1183) = 620.74$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-116.16	20.95	-0.12	-5.55	< 0.001***
ELL (0 = No, 1 = ELL)	-26.28	8.98	-0.07	-2.93	0.004**
FRL (0 = No, 1 = FRL)	-80.71	13.08	-0.14	-6.17	< 0.001***
Gender (0 = M, 1= F)	24.05	8.42	0.06	2.86	0.004**
Fall 2011 PASS Scaled Score	1.12	0.04	0.57	24.91	< 0.001***
Block 3: Demographics + Fall 2011 PASS Scaled Score + Phase Model Fit: $F(6, 1182) = 163.65$, $p < 0.001$, $R^2 = 0.454$ F Change $(1,1182) = 0.002$, $p = 0.969$					
IEP (0 = No, 1 = IEP)	-116.17	20.96	-0.12	-5.54	< 0.001***
ELL (0 = No, 1 = ELL)	-26.30	9.00	-0.07	-2.92	0.004**
FRL (0 = No, 1 = FRL)	-80.74	13.11	-0.14	-6.16	< 0.001***
Gender (0 = M, 1= F)	24.04	8.43	0.06	2.85	0.004**
Fall 2011 PASS Scaled Score	1.12	0.04	0.57	24.90	< 0.001***
Phase (0 = P2, 1 = P1)	-0.34	8.53	0.00	-0.04	0.969

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 26. Stanford Reading, HISD, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 NCE scores (*N* =1,189)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	688	457.2	199.20	450.4	501	441.5	188.88	450.8	0.00	0.969	0.00	50
Not IEP	660	464.6	196.37	458.3	478	449.2	184.95	458.0	0.00	0.971	0.00	50
IEP	28	282.0	187.92	282.6	23	280.6	202.08	279.8	0.00	0.952	0.01	51
Not ELL	354	505.8	203.12	491.3	221	485.2	201.54	508.4	1.96	0.162	-0.08	47
ELL	334	405.7	181.55	411.8	280	406.9	170.85	399.6	1.03	0.311	0.07	53
Not FRL	114	615.0	207.38	605.3	55	581.1	218.80	601.1	0.04	0.846	0.02	51
FRL	574	425.9	182.10	424.4	446	424.2	177.68	426.2	0.04	0.844	-0.01	50
Male	354	436.8	202.84	432.5	235	435.0	190.72	441.5	0.55	0.459	-0.05	48
Female	334	478.8	193.25	469.3	266	447.1	187.42	459.1	0.72	0.397	0.05	52

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Stanford Spring 2014 Results: Elementary Cohort Mathematics

For the 1,084 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 mathematics Stanford NCE scores (Block 3) explained 40% of the total variance (R^2) in students' 2013-2014 mathematics Stanford NCE scores (see Table 27). It also indicated a statistically significant difference in students' 2013-2014 Stanford NCE scores by Phase, favoring Phase 2 students, with a one percentage point increase in variance explained by the addition of student's Phase to the model (β = -0.08, t = -3.15, p = 0.002).

The overall ANCOVA analysis (see Table 28) revealed that there was a statistically significant difference in students' 2013-2014 mathematics Stanford NCE scores between Phase 1 and Phase 2 elementary students overall, and the effect size (g = -0.15), favoring Phase 2 students, was not substantively important according to WWC guidelines.

The ANCOVA analyses for the subgroup comparisons revealed that there were statistically significant differences between Phase 1 and Phase 2 in all subgroups except IEP and Female students. For the statistically significant effect sizes, only those associated with the ELL (g = -0.38) and Not FRL (g = 0.36) subgroups comparisons were substantively important, favoring Phase 1 students in Not FRL subgroup and Phase 2 students in ELL subgroup. In addition, the effect size for the FRL (g = -0.24) was nearly substantively important. Given the statistical and substantive baseline equivalence between Phase 1 and Phase 2 in the Not FRL subgroup, it appears that Phase 1 Not FRL elementary cohort students achieved advantages on the 2014 Stanford mathematics compared to their Phase 2 counterparts.

Table 27. Stanford Mathematics, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 NCE Scores (N = 1,084)

Source	*B*	*S.E.B.*	β	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1079) = 24.79$, $p < 0.001$, $R^2 = 0.084$ F Change $(4, 1079) = 24.79$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-181.25	27.85	-0.19	-6.51	< 0.001***
ELL (0 = No, 1 = ELL)	-10.48	12.37	-0.03	-0.85	0.397
FRL (0 = No, 1 = FRL)	-119.37	17.39	-0.21	-6.87	< 0.001***
Gender (0 = M, 1= F)	17.99	11.60	0.05	1.55	0.121
Block 2: Demographics + 2010-2011 Stanford mathematics NCE Score Model Fit: $F(5, 1078) = 138.64$, $p < 0.001$, $R^2 = 0.391$ F Change $(1,1078) = 544.12$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-51.45	23.38	-0.05	-2.20	0.028*
ELL (0 = No, 1 = ELL)	-110.27	10.96	-0.28	-10.06	< 0.001***
FRL (0 = No, 1 = FRL)	-73.55	14.31	-0.13	-5.14	< 0.001***
Gender (0 = M, 1= F)	18.34	9.46	0.05	1.94	0.053
2010-2011 Stanford mathematics NCE Score	5.54	0.24	0.62	23.33	< 0.001***
Block 3: Demographics + 2010-2011 Stanford mathematics NCE Score + Phase Model Fit: $F(6, 1077) = 118.13$, $p < 0.001$, $R^2 = 0.397$ F Change $(1, 1077) = 9.90$, $p = 0.002$					
IEP (0 = No, 1 = IEP)	-51.11	23.29	-0.05	-2.19	0.028*
ELL (0 = No, 1 = ELL)	-113.33	10.96	-0.29	-10.34	< 0.001***
FRL (0 = No, 1 = FRL)	-75.60	14.27	-0.14	-5.30	< 0.001***
Gender (0 = M, 1= F)	17.06	9.43	0.04	1.81	0.071
2010-2011 Stanford mathematics NCE Score	5.60	0.24	0.63	23.60	< 0.001***
Phase (0 = P2, 1 = P1)	-30.07	9.56	-0.08	-3.15	0.002**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 28. Stanford Mathematics, HISD, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 NCE scores (N = 1,084)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	616	567.7	201.14	556.4	468	571.7	194.59	586.5	9.90	0.002**	-0.15	44
Not IEP	589	574.9	196.84	563.7	446	580.2	186.59	595.0	10.39	0.001**	-0.16	44
IEP	27	409.4	231.17	417.4	22	400.3	268.25	390.5	0.24	0.630	0.11	54
Not ELL	314	605.1	202.34	599.1	199	564.3	199.25	573.9	3.98	0.047*	0.13	55
ELL	302	528.7	192.61	516.9	269	577.2	191.26	590.4	29.31	< 0.001***	-0.38	35
Not FRL	107	692.3	199.37	695.2	53	626.4	211.06	620.7	8.12	0.005**	0.36	64
FRL	509	541.5	191.63	531.5	415	564.7	191.53	577.0	19.98	< 0.001***	-0.24	41
Male	320	556.4	197.52	545.9	224	566.2	192.90	581.2	6.56	0.011*	-0.18	43
Female	296	579.9	204.60	567.9	244	576.8	196.39	591.3	3.09	0.079	-0.12	45

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Stanford Spring 2014 Results: Elementary Cohort Science

For the 1,189 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2011 PASS-Basic scaled scores (Block 3) explained 31% of the total variance (R^2) in students' 2013-2014 science Stanford NCE scores (see Table 29). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 science NCE scores (β = -0.02, t = -1.02, p = 0.307).

The overall ANCOVA analysis (see Table 30) revealed that there was neither a statistically significant nor a substantively important difference in students' 2013-2014 science Stanford NCE scores between Phase 1 and Phase 2 elementary cohort students overall. Consistent with the overall outcome, all subgroup ANCOVA analyses revealed neither statistically significant nor substantively important differences.

Table 29. Stanford Science, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 NCE Scores (*N* = 1,189)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1184) = 22.80$, $p < 0.001$, $R^2 = 0.072$ F Change $(4, 1184) = 22.80$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-156.30	28.01	-0.16	-5.58	< 0.001***
ELL (0 = No, 1 = ELL)	-16.43	12.03	-0.04	-1.37	0.172
FRL (0 = No, 1 = FRL)	-119.71	17.18	-0.21	-6.97	< 0.001***
Gender (0 = M, 1= F)	-10.48	11.38	-0.03	-0.92	0.357
Block 2: Demographics + Fall 2011 PASS Scaled Score Model Fit: $F(5, 1183) = 105.61$, $p < 0.001$, $R^2 = 0.309$ F Change $(1,1183) = 405.71$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-84.02	24.44	-0.08	-3.44	0.001**
ELL (0 = No, 1 = ELL)	13.05	10.48	0.03	1.24	0.214
FRL (0 = No, 1 = FRL)	-46.46	15.27	-0.08	-3.04	0.002**
Gender (0 = M, 1= F)	-11.93	9.82	-0.03	-1.21	0.225
Fall 2011 PASS Scaled Score	1.06	0.05	0.52	20.14	< 0.001***
Block 3: Demographics + Fall 2011 PASS Scaled Score + Phase Model Fit: $F(6, 1182) = 88.19$, $p < 0.001$, $R^2 = 0.309$ F Change $(1,1182) = 1.04$, $p = 0.307$					
IEP (0 = No, 1 = IEP)	-84.46	24.45	-0.08	-3.45	0.001**
ELL (0 = No, 1 = ELL)	12.50	10.50	0.03	1.19	0.234
FRL (0 = No, 1 = FRL)	-47.27	15.29	-0.08	-3.09	0.002**
Gender (0 = M, 1= F)	-12.40	9.84	-0.03	-1.26	0.208
Fall 2011 PASS Scaled Score	1.06	0.05	0.52	20.16	< 0.001***
Phase (0 = P2, 1 = P1)	-10.17	9.96	-0.02	-1.02	0.307

** $p < 0.01$; *** $p < 0.001$

Table 30. Stanford Science, HISD, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 NCE scores (*N* = 1,189)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	688	573.4	200.71	568.2	501	571.3	204.60	578.4	1.04	0.307	-0.05	48
Not IEP	660	579.2	198.26	574.3	478	577.9	199.02	584.6	1.03	0.309	-0.05	48
IEP	28	435.2	212.10	444.7	23	433.4	267.82	421.8	0.17	0.685	0.09	54
Not ELL	354	603.7	197.36	589.8	221	576.3	210.45	598.6	0.42	0.518	-0.04	48
ELL	334	541.3	199.51	546.7	280	567.3	200.15	560.7	0.92	0.338	-0.07	47
Not FRL	114	687.2	174.84	679.7	55	661.2	218.11	676.7	0.02	0.891	0.02	51
FRL	574	550.8	197.94	549.4	446	560.2	200.35	562.0	1.30	0.254	-0.06	47
Male	354	574.0	212.34	572.0	235	582.6	205.05	585.6	0.87	0.350	-0.06	47
Female	334	572.7	187.91	565.8	266	561.3	204.06	569.9	0.09	0.767	-0.02	49

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Stanford Spring 2014 Results: Middle School Cohort Reading

For the 291 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2012 PASS-Basic scaled scores (Block 3) explained 53% of the total variance (R^2) in students' 2013-2014 reading Stanford NCE scores (see Table 31). It also indicated a statistically significant difference in students' 2013-2014 Stanford NCE scores by Phase, favoring Phase 1, with a one percentage point increase in variance explained by the addition of student's Phase to the model ($\beta = 0.12$, $t = 2.80$, $p = 0.005$).

Based on the overall ANCOVA analysis for middle school cohort students (see Table 32), there was a statistically significant difference in students' 2013-2014 reading Stanford NCE scores between Phase 1 and Phase 2 students. The effect size ($g = 0.24$), favoring Phase 1 students, was nearly substantively important according to WWC guidelines, and indicated that the average Phase 1 student scored at the 59th percentile of the control group ($PR = 59$).

The ANCOVA analyses for the subgroup comparisons revealed that Phase 1 students statistically significantly outperformed their Phase 2 counterparts in the subgroups Not IEP ($g = 0.29$), FRL ($g = 0.21$) and Female ($g = 0.27$), and the effect sizes for the Not IEP and Female subgroups were substantively important according to WWC guidelines, indicating that the average Phase 1 Not IEP student scored at the 61st percentile of the control group ($PR = 61$), and the average Phase 1 Female student scored at the 61st percentile of the control group ($PR = 61$). For the Not IEP and Female subgroups, there were both statistical and substantive baseline equivalence between phase 1 and Phase 2. Therefore, it appears that that Phase 1 students in the Not IEP and Female subgroups achieved advantages on the 2014 Stanford reading compared to their respective Phase 2 counterparts. Although not statistically significant, the effect size for the Not FRL ($g = 0.61$) subgroup was substantively important, with the average Phase 1 Not FRL student scoring at the 73rd percentile of the control group ($PR = 73$). However, Phase 1 Not FRL students had a substantively important advantage on the pretest ($g = 0.36$). Therefore, the large effect sizes for the posttest (i.e., spring 2014 Stanford reading) could be a function of the large advantage they had at the pretest, and appears to indicate that Phase 1 students maintained their pretest advantage by spring 2014. In addition, the sample sizes, especially for the Phase 2 Not FRL students (Phase 1 $n = 16$; Phase 2 $n = 3$) were extremely small, and should therefore be treated with caution.

Table 31. Stanford Reading, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 NCE Scores (N = 291)

Source	*B*	*S.E.B.*	β	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 286) = 20.01$, $p < 0.001$, $R^2 = 0.219$ F Change (4, 286) = 20.01, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-242.45	28.79	-0.44	-8.42	< 0.001***
ELL (0 = No, 1 = ELL)	-9.04	21.28	-0.02	-0.42	0.671
FRL (0 = No, 1 = FRL)	-106.71	37.28	-0.15	-2.86	0.005**
Gender (0 = M, 1= F)	-1.36	18.59	0.00	-0.07	0.942
Block 2: Demographics + Spring 2012 PASS Scaled Score Model Fit: $F(5, 285) = 60.79$, $p < 0.001$, $R^2 = 0.516$ F Change (1, 285) = 175.16, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-137.21	24.05	-0.25	-5.71	< 0.001***
ELL (0 = No, 1 = ELL)	-13.01	16.78	-0.03	-0.78	0.439
FRL (0 = No, 1 = FRL)	-71.92	29.51	-0.10	-2.44	0.015*
Gender (0 = M, 1= F)	4.28	14.66	0.01	0.29	0.771
Spring 2012 PASS Scaled Score	0.85	0.06	0.58	13.23	< 0.001***
Block 3: Demographics + Spring 2012 PASS Scaled Score + Phase Model Fit: $F(6, 284) = 53.18$, $p < 0.001$, $R^2 = 0.529$ F Change (1, 284) = 7.83, $p = 0.005$					
IEP (0 = No, 1 = IEP)	-129.55	23.92	-0.24	-5.42	< 0.001***
ELL (0 = No, 1 = ELL)	-7.04	16.72	-0.02	-0.42	0.674
FRL (0 = No, 1 = FRL)	-57.28	29.63	-0.08	-1.93	0.054
Gender (0 = M, 1= F)	6.20	14.50	0.02	0.43	0.669
Spring 2012 PASS Scaled Score	0.86	0.06	0.59	13.52	< 0.001***
Phase (0 = P2, 1 = P1)	41.35	14.78	0.12	2.80	0.005**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 32. Stanford Reading, HISD, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 NCE scores (N = 291)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	p	g	PR
All	148	478.5	194.66	472.2	143	424.4	150.89	430.9	7.83	0.005**	0.24	59
Not IEP	135	503.1	181.44	501.9	122	454.9	125.32	456.2	8.76	0.003**	0.29	61
IEP	13	222.9	137.81	246.1	21	247.2	167.33	232.9	0.06	0.810	0.08	53
Not ELL	118	462.0	172.01	466.2	99	439.3	152.20	434.3	3.49	0.063	0.19	58
ELL	30	543.5	259.27	478.4	44	390.9	143.99	435.3	1.93	0.169	0.22	59
Not FRL	16	588.7	211.85	576.0	3	375.0	175.12	442.5	2.64	0.128	0.61	73
FRL	132	465.1	188.99	463.4	140	425.5	150.87	427.2	5.80	0.017*	0.21	58
Male	71	487.2	195.45	464.6	64	403.3	168.08	428.3	2.47	0.119	0.20	58
Female	77	470.4	194.86	478.5	79	441.6	134.03	433.7	5.10	0.025*	0.27	61

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$

Stanford Spring 2014 Results: Middle School Cohort Mathematics

For the 244 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 mathematics TAKS scaled scores (Block 3) explained 57% of the total variance (R^2) in students' 2013-2014 mathematics Stanford NCE scores (see

Table **33**). It also indicated a statistically significant difference in students' 2013-2014 Stanford NCE scores by Phase, favoring Phase 2, with a two percentage point increase in variance explained by the addition of student's Phase to the model (β = -0.15, t = -3.39, p = 0.001).

The overall ANCOVA analysis revealed that (see Table 34) there was a statistically significant difference in students' 2013-2014 mathematics Stanford NCE scores between Phase 1 and Phase 2 middle school students overall, and the effect size (g = 0.30) favoring Phase 1 students was substantively important according to WWC guidelines, with the average Phase 1 student scoring at the 62nd percentile of the control group (PR = 62). Given the statistical and substantive baseline equivalence between Phase 1 and Phase 2, it appears that, overall, Phase 1 middle school cohort students achieved advantages on the 2014 Stanford mathematics compared to their Phase 2 counterparts.

The ANCOVA analyses for the subgroup comparisons revealed that Phase 1 students statistically significantly outperformed their Phase 2 counterparts in the subgroups Not IEP, ELL, FRL, Male, and Female, and the effect sizes associated with these five subgroup comparisons, favoring Phase 1 students, were all substantively important:

- Not IEP (g = 0.30, PR = 62), with the average Phase 1 Not IEP student scoring at the 62nd percentile of the control group;
- ELL (g = 0.56, PR = 71), with the average Phase 1 ELL student scoring at the 71st percentile of the control group;
- FRL (g = 0.29, PR = 61), with the average Phase 1 FRL student scoring at the 61st percentile of the control group;
- Male (g = 0.38, PR = 65), with the average Phase 1 Male student scoring at the 65th percentile of the control group;
- Female (g = 0.25, PR = 60), with the average Phase 1 Female student scoring at the 60th percentile of the control group.

While not statistically significant, the effect size for the IEP (g = 0.35) and Not FRL (g = 0.60) reached the WWC threshold for substantive importance, and favored Phase 1 students, with the average Phase 1 IEP student scoring at the 64th percentile of the control group (PR = 64) and the average Phase 1 Not FRL student scoring at the 73rd percentile of the control group (PR = 73). In addition, while the comparison for the Not ELL subgroup was nearly statistically significant (p = 0.052), the associated effect size was not substantively important (g = 0.20). It should be noted that Phase 1 IEP and ELL students had substantively important advantages on the pretest (g = 0.63 and g = 0.44). Therefore, the large effect sizes for the posttest (i.e., spring 2014 Stanford mathematics) could be a function of the large advantage they had at the pretest, and appears to indicate that Phase 1 students maintained their pretest advantage by spring 2014. For subgroups Not IEP, FRL, Male and Female, there were both statistical and substantive baseline equivalences between Phase 1 and Phase 2; thus, it appears that Phase 1 students in each of these four subgroups achieved advantages on the 2014 Stanford mathematics compared to their Phase 2 counterparts. In addition, for the Not FRL subgroup, Phase 2 students had substantively higher baseline scores (g = -0.28). Therefore, it appears that Phase 1 Not FRL students were able to not only greatly reduce, but even to reverse the achievement gap present at the baseline. However, we should also note that the small sample sizes for the Not FRL subgroup, particularly for Phase 2 students (n = 2) would indicate that this outcome would not be representative of this subgroups' performance.

Table 33. Stanford Mathematics, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 NCE Scores (N = 244)

Source	*B*	*S.E.B.*	*ß*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 239) = 3.44$, $p = 0.009$, $R^2 = 0.054$ F Change (4, 239) = 3.44, $p = 0.009$					
IEP (0 = No, 1 = IEP)	-134.47	69.68	-0.12	-1.93	0.055
ELL (0 = No, 1 = ELL)	21.35	26.98	0.05	0.79	0.430
FRL (0 = No, 1 = FRL)	-135.99	44.09	-0.19	-3.08	0.002**
Gender (0 = M, 1= F)	-7.99	23.30	-0.02	-0.34	0.732
Block 2: Demographics + 2010-2011 TAKS mathematics Scaled Scores Model Fit: $F(5, 238) = 58.71$, $p < 0.001$, $R^2 = 0.552$ F Change (1, 238) = 264.60, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-17.41	48.59	-0.02	-0.36	0.720
ELL (0 = No, 1 = ELL)	42.96	18.66	0.10	2.30	0.022*
FRL (0 = No, 1 = FRL)	-89.20	30.54	-0.13	-2.92	0.004**
Gender (0 = M, 1= F)	5.30	16.09	0.01	0.33	0.742
2010-2011 TAKS mathematics Scaled Scores	1.33	0.08	0.72	16.27	< 0.001***
Block 3: Demographics + 2010-2011 TAKS mathematics Scaled Scores + Phase Model Fit: $F(6, 237) = 52.99$, $p < 0.001$, $R^2 = 0.573$ F Change (1, 237) = 11.48, $p = 0.001$					
IEP (0 = No, 1 = IEP)	-7.78	47.64	-0.01	-0.16	0.870
ELL (0 = No, 1 = ELL)	49.69	18.37	0.12	2.71	0.007**
FRL (0 = No, 1 = FRL)	-68.86	30.49	-0.10	-2.26	0.025*
Gender (0 = M, 1= F)	9.25	15.79	0.03	0.59	0.558
2010-2011 TAKS mathematics Scaled Scores	1.32	0.08	0.71	16.51	< 0.001***
Phase (0 = P2, 1 = P1)	-54.39	16.06	-0.15	-3.39	0.001**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 34. Stanford Mathematics, HISD, Spring 2014: S Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 NCE scores (*N* = 244)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	131	607.9	194.63	599.2	113	534.7	161.53	544.8	11.48	0.001**	0.30	62
Not IEP	128	609.2	194.01	602.8	109	540.6	159.51	548.2	11.2	0.001**	0.30	62
IEP	3	552.0	259.47	496.9	4	372.8	147.99	414.1	0.84	0.455	0.35	64
Not ELL	104	585.5	183.59	582.6	80	544.4	154.98	548.1	3.82	0.052	0.20	58
ELL	27	694.2	214.77	653.9	33	511.2	176.70	544.2	8.54	0.005**	0.56	71
Not FRL	16	707.1	194.12	709.5	2	608.5	108.19	589.8	0.94	0.351	0.60	73
FRL	115	594.1	191.48	589.4	111	533.4	162.35	538.2	9.98	0.002**	0.29	61
Male	64	616.7	193.24	608.7	48	526.2	176.70	536.9	7.99	0.006**	0.38	65
Female	67	599.5	197.03	592.4	65	540.9	150.45	548.3	4.48	0.036*	0.25	60

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$

Stanford Spring 2014 Results: Middle School Cohort Science

For the 291 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2012 PASS-Basic scaled scores (Block 3) explained 54% of the total variance (R^2) in students' 2013-2014 science Stanford NCE scores (see Table 35). It also indicated a statistically significant difference in students' 2013-2014 Stanford NCE scores by Phase, favoring Phase 2, with a one percentage point increase in variance explained by the addition of student's Phase to the model (β = -0.12, t = -2.79, p = 0.006).

The overall ANCOVA analysis (see Table 36) revealed that, overall, Phase 2 middle school cohort students statistically significantly outperformed their Phase 1 counterparts in 2013-2014 Stanford science, and the effect size (g = -0.23) was very close to the WWC threshold for statistical significance, indicating the average Phase 1 student scored at the 41st percentile of the control group (PR = 41).

The ANCOVA analyses for the subgroup comparisons revealed that Phase 2 students statistically significantly outperformed Phase 1 students for the Not IEP (g = -0.23), ELL (g = -0.45), FRL (g = -0.25), and Male (g = -0.29) subgroups. The effect sizes for the ELL, FRL, and Male subgroups comparisons were substantively important, and was nearly substantively for the Not IEP subgroup, indicating that the average Phase 1 ELL student scored at the 33rd percentile of the control group (PR = 33), the average Phase 1 FRL student scored at the 40th percentile of the control group (PR = 40), and the average Phase 1 Male student scored at the 38th percentile of the control group (PR = 38). Although not statistically significant, the effect size associated with the IEP (g = -0.56) and Not FRL (g = 0.37) subgroup comparisons were substantively important, indicating that the average Phase 1 IEP student scored at the 29th percentile of the control group (PR = 29) and the average Phase 1 Not FRL student scored at the 65th percentile of the control group (PR = 65). Note that for the Not FRL subgroup, Phase 1 students had a substantively important advantage on the pretest (g = 0.36). Thus, the large effect size for the posttest (i.e., spring 2014 Stanford science) could be a function of the large advantage they had at the pretest, and appears to indicate that Phase 1 Not FRL students maintained their pretest advantage by spring 2014. The comparisons for the Not ELL and Female subgroups were neither statistically significant nor substantively important.

Table 35. Stanford Science, HISD, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 NCE Scores ($N = 291$)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 286) = 12.56$, $p < 0.001$, $R^2 = 0.149$ F Change $(4, 286) = 12.56$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-213.51	31.86	-0.37	-6.70	< 0.001***
ELL (0 = No, 1 = ELL)	3.30	23.55	0.01	0.14	0.889
FRL (0 = No, 1 = FRL)	-79.84	41.26	-0.11	-1.94	0.054
Gender (0 = M, 1= F)	-27.24	20.57	-0.07	-1.32	0.187
Block 2: Demographics + Spring 2012 Test Score Scaled Model Fit: $F(5, 285) = 63.55$, $p < 0.001$, $R^2 = 0.527$ F Change $(1, 285) = 227.68$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-87.71	25.21	-0.15	-3.48	0.001**
ELL (0 = No, 1 = ELL)	-1.45	17.59	0.00	-0.08	0.934
FRL (0 = No, 1 = FRL)	-38.25	30.94	-0.05	-1.24	0.217
Gender (0 = M, 1= F)	-20.49	15.37	-0.05	-1.33	0.184
Spring 2012 Test Score Scaled	1.02	0.07	0.65	15.09	< 0.001***
Block 3: Demographics + Spring 2012 Test Score Scaled + Phase Model Fit: $F(6, 284) = 55.53$, $p < 0.001$, $R^2 = 0.540$ F Change $(1, 284) = 7.81$, $p = 0.006$					
IEP (0 = No, 1 = IEP)	-95.72	25.08	-0.16	-3.82	< 0.001***
ELL (0 = No, 1 = ELL)	-7.70	17.53	-0.02	-0.44	0.661
FRL (0 = No, 1 = FRL)	-53.57	31.06	-0.07	-1.72	0.086
Gender (0 = M, 1= F)	-22.51	15.21	-0.06	-1.48	0.140
Spring 2012 Test Score Scaled	1.01	0.07	0.65	15.11	< 0.001***
Phase (0 = P2, 1 = P1)	-43.29	15.49	-0.12	-2.79	0.006**

** $p < 0.01$; *** $p < 0.001$

Table 36. Stanford Science, HISD, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 NCE scores ($N = 291$)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	148	561.4	211.10	555.9	143	593.4	157.40	599.2	7.81	0.006**	-0.23	41
Not IEP	135	585.6	200.52	583.1	122	619.9	139.85	622.6	5.95	0.015*	-0.23	41
IEP	13	310.7	149.09	332.8	21	439.7	168.82	426.0	2.58	0.120	-0.56	29
Not ELL	118	551.8	196.45	558.1	99	599.0	159.73	591.4	3.37	0.068	-0.18	43
ELL	30	599.5	261.36	533.8	44	580.9	153.09	625.6	7.79	0.007**	-0.45	33
Not FRL	16	675.1	178.01	667.5	3	552.7	257.32	593.6	0.68	0.425	0.37	65
FRL	132	547.6	211.21	547.5	140	594.3	155.95	594.4	8.73	0.003**	-0.25	40
Male	71	583.6	215.48	560.0	64	591.8	171.61	617.9	6.17	0.014*	-0.29	38
Female	77	541.0	206.27	550.2	79	594.7	145.98	585.7	2.73	0.101	-0.20	42

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$

New Mexico Region: Results for Spring 2014 State Assessments

New Mexico
Spring 2014 Standards Based Assessment Tests (SBA)
Key Findings for Phase 1

For all students combined (the “All” group) and the specified subgroups in the New Mexico region, the following outcomes favoring Phase 1 students were found on the Spring 2014 SBA in reading.

IEP

- <u>Elementary Cohort</u>: Phase 1 had a substantively higher adjusted mean score than Phase 2 in Spring 2014 (*g* = 0.27).

ELL

- <u>Middle School Cohort</u>: Phase 1 had a substantively higher adjusted mean score than Phase 2 in Spring 2014 (*g* = 0.30). However, the sample size for Phase 1 (*n* = 23) was small.

New Mexico
Standards Based Assessment Tests: Spring 2014

New Mexico Standards Based Assessment (SBA) results from the Spring 2010 or Spring 2011 (baseline or pre-intervention, where available) and Spring 2014 (third posttest) administrations are currently available and are reported below. It should be noted that as the PASS assessment is better aligned with and more sensitive to changes in program outcomes (i.e., inquiry-based science instruction and knowledge/application), results from the state assessments should be interpreted judiciously when being used to evaluate program impacts.

New Mexico: Elementary and Middle School Cohorts Standards Based Assessment (SBA) Spring 2014 Analyses

There were a total of 826 elementary cohort students in Phase 1 (n = 509) and Phase 2 (n = 317) schools and 579 middle school cohort students in Phase 1 (n = 467) and Phase 2 (n = 112) schools included in the analysis. To be included in the analysis, a student had to meet two criteria: 1) a student had to have scores on the multiple choice sections of PASS in both Fall 2011 and Spring 2014, and 2) a student had to take the Spring 2014 reading SBA and the selected baseline achievement assessment. With respect to the 826 elementary cohort students and the 579 middle school cohort students, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students within cohort grade levels differed by Phase in their performance on 2013-2014 New Mexico SBA reading[2] scaled scores. In addition to these regressions, a second set of analyses (ANCOVA) intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (g) were also conducted on the outcomes for all students by Phase within cohort grade level, as well as for subgroups of these same students, categorized by their IEP (Special Education) status, ELL (English Language Learner) status, Economically Disadvantaged (FRL) status, and Gender. As the analyses were all exploratory in nature, no corrections were made for multiple comparisons.

In the selection of the baseline achievement test, four major factors were considered: (1) the number of students available for analysis; (2) the correlation between the baseline and current test scores; (3) whether or not the ANCOVA assumption of homogeneity of variance was met; and (4) independent t-test results (i.e., whether or not there was a non-significant difference in the baseline achievement between Phase 1 and Phase 2 students overall and by subgroups).

It should be noted that because students in the elementary cohort do not have baseline (Spring 2011) SBA test scores available in reading, the Fall 2011 PASS scaled score was used as the prior-achievement measure for the analyses, with the correlation between the Spring 2014 reading SBA scaled score and the Fall 2011 PASS scaled score being moderately strong and statistically significant ($r = 0.56$, $p < 0.001$). The Spring 2011 5th grade reading SBA was used as the pretest for the middle school cohort reading analyses, with the correlation between the Spring 2011 and Spring 2014 SBA reading scaled scores being high and statistically significant ($r = 0.77$, $p < 0.001$).

To determine baseline achievement score equivalence between Phase 1 and Phase 2 students included in the present analysis, a series of independent t-tests was conducted for all elementary and middle

[2] The science test in New Mexico is only given in 4th and 7th grades, therefore science scores are not available for spring 2014 (when the cohorts were in 5th and 8th grades). Thus, only reading scores are analyzed in the current report.

school cohort students in the aggregate as well as for subgroups of these students by their Special Education (IEP) status, English language learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender. In addition, an effect size was also calculated as a measure of baseline equivalence.

As an indicator of the impact or "practical significance" of the treatment, the "effect size" (calculated as Hedges's *g*) is a descriptive statistic that indicates the magnitude of the difference (in standard deviation units) between two measures. For example, a positive effect size would indicate a higher (i.e., better) Phase 1 mean, while a negative effect size would indicate a higher (i.e., better) Phase 2 mean. Based on guidelines from the What Works Clearinghouse (WWC), a unit within the research division of the U.S. Department of Education, an effect size of +/- 0.25 is considered to be "substantively important" (What Works Clearinghouse, 2014).

With respect to the elementary cohort (Table 37), no statistically significant differences by phase in the baseline achievement levels were found for students in either the aggregate (the "All" group) or any subgroups except the Not FRL subgroup (t (433) = 2.28, p = 0.02, g = 0.22, *PR* = 63), but the effect size was not substantively important.

Table 37. SBA Reading Achievement Spring 2014 Scaled Scores, New Mexico: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – Fall 2011 PASS-B Scaled Scores (*N* = 826)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort									
All	509	329.7	105.80	317	317.8	112.50	1.53	0.11	54
Not IEP	445	336.4	104.90	273	324	112.60	1.49	0.11	55
IEP	64	282.9	100.70	44	279.3	104.90	0.18	0.03	51
Not ELL	444	339.5	103.30	280	325.6	111.60	1.71	0.13	55
ELL	65	262.6	98.69	37	259	102.70	0.17	0.04	51
Not FRL	267	365.7	98.37	168	342.2	113.80	2.28*	0.22	59
FRL	242	289.8	99.43	149	290.3	104.70	-0.04	0.00	50
Male	258	334.1	112.80	158	319.7	114.40	1.26	0.13	55
Female	251	325.1	98.15	159	316	110.90	0.87	0.09	54

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$

With respect to students in the middle school cohort (Table 38), there were statistically significant and substantively important differences in baseline achievement by phase in the aggregate (t (577) = 4.14, p < 0.001, g = 0.44, *PR* = 67) and four subgroups (Not IEP: t (511) = 3.06, p = 0.002, g = 0.34, *PR* = 63; FRL: t (397) = 2.65, p = 0.008, g = 0.30, *PR* = 62; Male: t (273) = 2.67, p = 0.008, g = 0.40, *PR* = 66; Female: t (302) = 3.25, p = 0.001, g = 0.48, *PR* = 68), with Phase 1 students being favored in each case. In addition, the effect size associated with the difference between Phase 1 and Phase 2 in the IEP subgroup was very close to the WWC threshold for substantive difference, g = 0.24. Therefore, with respect to students in the middle school cohort, the outcomes should be interpreted in light of the substantively important difference in baseline achievement between Phase 1 and Phase 2 students in the aggregate and the following four subgroups: Not IEP, FRL, Male and Female subgroups. Special caution should be exercised for the middle school cohort given that in all groups except ELL subgroup the sample sizes for Phase 1 were at least twice that of Phase 2.

Table 38. SBA Reading Achievement Spring 2014 Scaled Scores, New Mexico: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – Spring 2011 Reading Scaled Scores (*N* = 579)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Middle School Cohort									
All	467	541.7	9.20	112	537.5	10.63	4.14***	0.44	67
Not IEP	422	543.1	7.79	91	540.4	8.46	3.06**	0.34	63
IEP	45	527.8	9.97	21	525.3	10.60	0.92	0.24	60
Not ELL	444	542.3	8.84	80	540.7	8.90	1.44	0.18	57
ELL	23	530.2	8.60	32	529.6	10.56	0.22	0.06	52
Not FRL	174	544.3	7.42	6	545.0	5.97	-0.24	-0.09	46
FRL	293	540.1	9.81	106	537.1	10.70	2.65**	0.30	62
Male	220	539.4	9.79	55	535.3	11.58	2.67**	0.40	66
Female	247	543.7	8.16	57	539.7	9.23	3.25**	0.48	68

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

** $p < 0.01$; *** $p < 0.001$

Elementary Cohort Reading SBA Spring 2014 Results

For the 826 students in the Elementary Cohort, the hierarchical multiple regression that controlled for student's demographic characteristics (IEP, ELL, FRL, and Gender) and their Fall 2011 PASS-Basic scaled scores (Block 3) explained 43.8% of the total variance (R^2) in students' 2014 Spring SBA reading scores (see Table 39). The addition of the student's Phase (i.e., Phase 1 or Phase 2) to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of Spring 2014 reading SBA scaled scores (β = 0.01, t = 0.18, p = 0.860).

The overall ANCOVA analysis (see Table 40) revealed that there was neither a statistically significant nor substantively important difference in students' 2013-2014 reading SBA scaled scores between Phase 1 and Phase 2 elementary cohort students overall.

According to the subgroup ANCOVA analyses, Phase 1 students scored higher in the subgroups IEP, ELL, Not FRL, and Male, while Phase 2 students scored higher in each of the other subgroups. However, none of the subgroup differences was statistically significant. Although not statistically significant, the effect size associated with the IEP subgroup comparison (g = 0.27) was substantively important, favoring Phase 1 students. Specifically, the average IEP Phase 1 student scored at the 61st percentile of the IEP Phase 2 students. In addition, the effect size associated with the ELL subgroup comparison (g = 0.24) was nearly substantively important given the WWC threshold for substantive difference (i.e., $g \geq 0.25$). The effect sizes for all other subgroup comparisons were not substantively important, ranging from -0.13 to 0.14.

Table 39. SBA Reading, New Mexico, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 Scaled Scores (N = 826)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 821) = 75.24$, $p < 0.001$, $R^2 = 0.268$ F Change $(4,821) = 75.24$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-9.69	0.96	-0.31	-10.05	< 0.001***
ELL (0 = No, 1 = ELL)	-6.59	1.00	-0.20	-6.60	< 0.001***
FRL (0 = No, 1 = FRL)	-6.73	0.66	-0.31	-10.20	< 0.001***
Gender (0 = M, 1= F)	2.61	0.65	0.12	4.03	< 0.001***
Block 2: Demographics + Fall Score Model Fit: $F(5, 820) = 127.93$, $p < 0.001$, $R^2 = 0.438$ F Change $(1,820) = 248.08$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-6.94	0.86	-0.22	-8.05	< 0.001***
ELL (0 = No, 1 = ELL)	-4.22	0.89	-0.13	-4.75	< 0.001***
FRL (0 = No, 1 = FRL)	-3.99	0.60	-0.19	-6.61	< 0.001***
Gender (0 = M, 1= F)	3.19	0.57	0.15	5.60	< 0.001***
Fall 2011 Test Score Scaled	0.04	0.00	0.45	15.75	< 0.001***
Block 3: Demographics + Fall Score + Phase **Model Fit: $F(6, 819) = 106.48$, $p < 0.001$, $R^2 = 0.438$** **F Change $(1,819) = 0.03$, $p = 0.860$**					
IEP (0 = No, 1 = IEP)	-6.94	0.86	-0.22	-8.04	< 0.001***
ELL (0 = No, 1 = ELL)	-4.22	0.89	-0.13	-4.75	< 0.001***
FRL (0 = No, 1 = FRL)	-3.99	0.60	-0.19	-6.61	< 0.001***
Gender (0 = M, 1= F)	3.19	0.57	0.15	5.60	< 0.001***
Fall 2011 Test Score Scaled	0.04	0.00	0.45	15.71	< 0.001***
Phase (0 = P2, 1 = P1)	**0.10**	**0.58**	**0.01**	**0.18**	**0.860**

*** $p < 0.001$

Table 40. SBA Reading, New Mexico, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 826)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
New Mexico: Elementary Cohort												
All	509	543.8	10.75	543.6	317	543.2	10.67	543.5	0.03	0.86	0.01	50
Not IEP	445	544.7	10.18	544.6	273	544.8	9.17	545.0	0.55	0.46	-0.04	48
IEP	64	537.1	12.24	536.8	44	533.1	13.56	533.4	2.83	0.10	0.27	61
Not ELL	444	544.8	10.26	544.5	280	544.3	10.12	544.8	0.14	0.71	-0.02	49
ELL	65	536.9	11.57	536.9	37	534.2	10.64	534.2	2.35	0.13	0.24	59
Not FRL	267	547.8	9.32	547.4	168	545.5	10.45	546.1	2.85	0.09	0.14	55
FRL	242	539.3	10.49	539.3	149	540.5	10.33	540.6	2.56	0.11	-0.13	45
Male	258	542.6	10.88	542.3	158	540.3	10.77	540.9	2.74	0.10	0.13	55
Female	251	544.9	10.52	544.9	159	546.0	9.82	546.1	2.31	0.13	-0.12	45

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Middle School Cohort Reading SBA Spring 2014 Results

For the 579 students in the middle school cohort, the hierarchical multiple regression that controlled for student's demographic characteristics (IEP, ELL, FRL, and Gender) and their Spring 2011 reading SBA scaled score (Block 3) explained 60% of the total variance (R^2) in students' 2014 Spring SBA reading scores (see Table 41). The addition of the student's Phase) to the model did not add to the percentage of variance explained, and there was no statistically significant difference in 2013-2014 reading scaled scores, on average, between Phase 1 and Phase 2 students taking into account all of the other variables in the previous blocks ($\beta = 0.05$, $t = 1.67$, $p = 0.096$).

The overall ANCOVA analysis (see Table 42) revealed that there was neither statistically significant nor substantively important difference in students' 2013-2014 reading SBA scaled scores between Phase 1 and Phase 2 middle school cohort students overall. Consistent with the overall outcome, all subgroup ANCOVA analyses revealed no statistically significant differences. However, the effect size associated with ELL subgroup comparison ($g = 0.30$) was found substantively important, with the average ELL Phase 1 students scoring at the 62nd percentile of the ELL Phase 2 students. The effect sizes for all other subgroup comparisons were not substantively important, ranging from -0.09 to 0.16.

Table 41. SBA Reading, New Mexico, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 Scaled Scores (N = 579)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 574) = 56.27$, $p < 0.001$, $R^2 = 0.282$ F Change (4574) = 56.27, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-11.89	1.20	-0.37	-9.87	< 0.001***
ELL (0 = No, 1 = ELL)	-5.96	1.31	-0.17	-4.55	< 0.001***
FRL (0 = No, 1 = FRL)	-3.87	0.80	-0.17	-4.82	< 0.001***
Gender (0 = M, 1= F)	2.39	0.75	0.12	3.20	0.001**
Block 2: Demographics + Fall Score Model Fit: $F(5, 573) = 168.89$, $p < 0.001$, $R^2 = 0.596$ F Change (1,1557) = 600.06, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-2.20	1.01	-0.07	-2.16	0.031*
ELL (0 = No, 1 = ELL)	-0.61	1.02	-0.02	-0.60	0.551
FRL (0 = No, 1 = FRL)	-1.26	0.62	-0.06	-2.05	0.041*
Gender (0 = M, 1= F)	0.74	0.57	0.04	1.31	0.191
2010 – 2011 reading Scaled Score	0.75	0.04	0.70	21.10	< 0.001
Block 3: Demographics + Fall Score + Phase **Model Fit: $F(6, 572) = 141.64$, $p < 0.001$, $R^2 = 0.598$** **F Change (1,572) = 2.785, $p = 0.096$**					
IEP (0 = No, 1 = IEP)	-2.15	1.01	-0.07	-2.12	0.034*
ELL (0 = No, 1 = ELL)	-0.15	1.05	0.00	-0.15	0.885
FRL (0 = No, 1 = FRL)	-1.03	0.63	-0.05	-1.64	0.103
Gender (0 = M, 1= F)	0.76	0.57	0.04	1.35	0.178
2010 – 2011 reading Scaled Score	0.75	0.04	0.70	21.12	< 0.001***
Phase (0 = P2, 1 = P1)	**1.25**	**0.75**	**0.05**	**1.67**	**0.096**

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 42. SBA Reading, New Mexico, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 579)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
New Mexico: Middle School Cohort												
All	467	845.2	10.08	844.5	112	840.3	10.27	843.3	2.79	0.10	0.12	55
Not IEP	422	846.6	8.86	846.2	91	842.6	8.96	844.7	3.13	0.08	0.16	56
IEP	45	832.0	11.30	831.7	21	830.2	9.59	830.9	0.15	0.70	0.08	53
Not ELL	444	845.7	9.98	845.4	80	843.1	9.19	844.5	1.41	0.24	0.10	54
ELL	23	836.0	7.62	835.9	32	833.2	9.42	833.3	2.00	0.16	0.30	62
Not FRL	174	847.8	9.20	847.8	6	849.0	6.90	848.6	0.08	0.77	-0.09	47
FRL	293	843.7	10.29	843.0	106	839.8	10.23	841.6	3.39	0.07	0.14	56
Male	220	842.9	9.96	842.2	55	838.3	11.51	841.1	1.05	0.31	0.11	54
Female	247	847.2	9.78	846.5	57	842.2	8.57	845.2	1.62	0.21	0.14	56

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$.

North Carolina Region: Results for Spring 2014 State Assessments

North Carolina
Spring 2014 Standards End-Of-Grade (EOG)
Key Findings for Phase 1

For all students combined (the "All" group) and the specified subgroups in the North Carolina region, the following outcomes favoring Phase 1 students were found on the Spring 2014 EOG.

ELL

- Middle School Cohort in science: Phase 1 had a substantively higher adjusted mean score than Phase 2 in Spring 2014 (*g* = 0.26).

- Middle School Cohort in mathematics: While Phase 2 ELL students had a substantively important advantage on the pretest (*g* = -0.42), Phase 1 ELL students had a higher adjusted mean score than Phase 2 in Spring 2014 (*g* = 0.11). Therefore, it appears that Phase 1 ELL students were able to not only greatly reduce, but even to reverse the achievement gap present at the baseline.

North Carolina

End-of-Grade Tests: Spring 2014

North Carolina End-Of-Grade (EOG) results from Spring 2011 (baseline or pre-intervention, where available) and Spring 2014 (third posttest) administrations are currently available and are reported below. It should be noted that as the PASS assessment is better aligned with and more sensitive to changes in program outcomes (i.e., inquiry-based science instruction and knowledge/application), results from the state assessments should be interpreted judiciously when being used to evaluate program impacts.

North Carolina: Elementary and Middle School Cohorts End-Of-Grade (EOG) Spring 2014 Analyses

There were a total of 1,847 elementary cohort students in Phase 1 (n = 886) and Phase 2 (n = 961) schools and 1,410 middle school cohort students in Phase 1 (n = 522) and Phase 2 (n = 888) schools for the analysis of the EOG test in reading, a total of 1,846 elementary cohort students in Phase 1 (n = 886) and Phase 2 (n = 960) schools and 1,410 middle school cohort students in Phase 1 (n = 522) and Phase 2 (n = 888) schools for the analysis of the EOG test in mathematics, and 1,847 elementary cohort students in Phase 1 (n = 886) and Phase 2 (n = 961) schools and 1,409 middle school cohort students in Phase 1 (n = 522) and Phase 2 (n = 887) schools for the analysis of the EOG test in science.

To be included in the analysis, a student had to meet two criteria: 1) a student had to have scores on the multiple choice section of PASS in both Fall 2011 and Spring 2014, and 2) a student had to take the Spring 2014 EOG in reading, mathematics, or science along with the selected baseline achievement assessment. With respect to the students included in the analysis, hierarchical or “block entry” multiple regressions were conducted to determine whether groups of students within cohort grade levels differed by Phase in their performance on 2013-2014 EOG reading, mathematics and science scaled scores. In addition to these regressions, a second set of analyses (ANCOVA) intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (g) were also conducted on the outcomes for all students by Phase within cohort grade level, as well as for subgroups of these same students, categorized by their IEP (Special Education) status, ELL (English Language Learner) status, Economically Disadvantaged (FRL) status, and Gender. As the analyses were all exploratory in nature, no corrections were made for multiple comparisons.

In the selection of the baseline achievement test, four major factors were considered: (1) the number of students available for analysis; (2) the correlation between the baseline and current test scores; (3) whether or not the ANCOVA assumption of homogeneity of variance was met; and (4) independent t-test results (i.e., whether or not there was a non-significant difference in the baseline achievement between Phase 1 and Phase 2 students overall and by subgroups).

It should be noted that because students in the elementary cohort do not have baseline (Spring 2011) EOG test scores in either reading, mathematics or science available, the Fall 2011 PASS scaled score was used as the prior-achievement measure for the analyses, with the correlation between the Spring 2014 reading EOG scaled score and the Fall 2011 PASS scaled score being low but statistically significant (r = 0.41, p < 0.001). There was little if any correlation between the Spring 2014 mathematics EOG scaled score and the Fall 2011 PASS scaled score, but the correlation was statistically significant (r = 0.28, p < 0.001). Finally, there was also little if any correlation between the Fall 2011 PASS scaled score and the 2013-2014 science EOG scaled score, but the correlation was statistically significant (r = 0.15, p < 0.001).

The Spring 2011 5th grade reading, mathematics, and science EOG scaled scores were used as the pretests for the middle school cohort Spring 2014 EOG reading, mathematics and science analyses, respectively. The correlation between the 2010-2011 reading EOG scaled score and the 2013-2014 reading EOG scaled score was high and statistically significant (r = 0.83, p < 0.001). Correlation between the 2010-2011 mathematics EOG scaled score and the 2013-2014 mathematics EOG scaled score was also high and statistically significant (r = 0.80, p < 0.001). Meanwhile, there was little if any correlation between the 2010-2011 science EOG scaled score and the 2013-2014 science EOG scaled score, but the correlation was statistically significant (r = 0.26, p < 0.001).

To determine baseline equivalence in achievement between Phase 1 and Phase 2 students included in the present analysis, a series of independent t-tests was conducted for all elementary and middle school cohort students in the aggregate as well as for subgroups of these students by their Special Education (IEP) status, English language learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender. In addition, an effect size was also calculated as a measure of baseline equivalence.

As an indicator of the impact or “practical significance” of the treatment, the “effect size” (calculated as Hedges’s g) is a descriptive statistic that indicates the magnitude of the difference (in standard deviation units) between two measures. For example, a positive effect size would indicate a higher (i.e., better) Phase 1 mean, while a negative effect size would indicate a higher (i.e., better) Phase 2 mean. Based on guidelines from the What Works Clearinghouse (WWC), a unit within the research division of the U.S. Department of Education, an effect size of +/- 0.25 is considered to be “substantively important” (i.e., educationally meaningful) (What Works Clearinghouse, 2014).

With respect to the elementary cohort in reading(Table 43), mathematics (Table 44), and science (Table 45), neither statistically significant nor substantively important differences by phase in the baseline achievement levels were found for students in either the aggregate (the “All” group) or any subgroups.

Table 43. EOG Reading Achievement Spring 2014 Scaled Scores, North Carolina: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – Fall 2011 PASS-B Scaled Scores (N = 1,847)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort									
All	886	326.4	93.21	961	326.5	92.83	-0.03	0.00	50
Not IEP	795	332.1	92.55	877	332.9	90.59	-0.19	-0.01	50
IEP	91	276.7	84.20	84	259.9	90.19	1.28	0.19	58
Not ELL	821	331.2	92.10	864	332.7	93.32	-0.34	-0.02	49
ELL	65	265.5	85.89	97	271.4	66.85	-0.49	-0.08	47
Not FRL	478	348.8	88.55	501	354.1	89.83	-0.92	-0.06	48
FRL	408	300.1	91.75	460	296.6	86.64	0.58	0.04	52
Male	446	330.0	93.52	503	327.6	94.80	0.39	0.03	51
Female	440	322.7	92.85	458	325.4	90.71	-0.43	-0.03	49

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (g). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Table 44. EOG Mathematics Achievement Spring 2014 Scaled Scores, North Carolina: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – Fall 2011 PASS-B Scaled Scores (*N* = 1,846)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort									
All	886	326.4	93.21	960	326.6	92.82	-0.06	0.00	50
Not IEP	795	332.1	92.55	877	332.9	90.59	-0.19	-0.01	50
IEP	91	276.7	84.20	83	260.3	90.67	1.24	0.19	57
Not ELL	821	331.2	92.10	863	332.9	93.31	-0.36	-0.02	49
ELL	65	265.5	85.89	97	271.4	66.85	-0.49	-0.08	47
Not FRL	478	348.8	88.55	501	354.1	89.83	-0.92	-0.06	48
FRL	408	300.1	91.75	459	296.7	86.67	0.56	0.04	52
Male	446	330.0	93.52	503	327.6	94.80	0.39	0.03	51
Female	440	322.7	92.85	457	325.6	90.70	-0.47	-0.03	49

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Table 45. EOG Science Achievement Spring 2014 Scaled Scores, North Carolina: Baseline Subgroup Mean Comparison of Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – Fall 2011 PASS-B Scaled Scores (*N* = 1,847)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort									
All	886	326.4	93.21	961	326.5	92.83	-0.03	0.00	50
Not IEP	795	332.1	92.55	877	332.9	90.59	-0.19	-0.01	50
IEP	91	276.7	84.20	84	259.9	90.19	1.28	0.19	58
Not ELL	821	331.2	92.10	864	332.7	93.32	-0.34	-0.02	49
ELL	65	265.5	85.89	97	271.4	66.85	-0.49	-0.08	47
Not FRL	478	348.8	88.55	501	354.1	89.83	-0.92	-0.06	48
FRL	408	300.1	91.75	460	296.6	86.64	0.58	0.04	52
Male	446	330	93.52	503	327.6	94.80	0.39	0.03	51
Female	440	322.7	92.85	458	325.4	90.71	-0.43	-0.03	49

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

For the middle school cohort in reading (see Table 46), neither statistically significant nor substantively important differences by phase in the baseline achievement levels were found for students in either the aggregate (the "All" group) or any subgroups. With respect to students in the middle school cohort in mathematics (Table 47), while no statistically significant differences by phase in the baseline achievement levels were found for students in either the aggregate or any subgroups, the effect size associated with the difference in the ELL subgroup (t (114) = -1.28 p = 0.203, g = -0.42, PR = 34) met the WWC threshold for substantive importance, favoring Phase 2 students. No other differences were substantively important, with the effect sizes ranging from -0.16 to 0.06. For the middle school cohort in science (see Table 48), like in reading, neither statistically significant nor substantively important differences by phase in the baseline achievement levels were found for students in either the aggregate (the "All" group) or any subgroups.

Therefore, the outcomes should be interpreted cautiously for the ELL middle school cohort students in mathematics in light of the substantively important difference in baseline achievement between Phase 1 and Phase 2 students (favoring Phase 2). Baseline achievement score equivalence (both statistical and substantive) between Phase 1 and Phase 2 students was established for all other groups in both elementary and middle school cohorts.

Table 46. EOG Reading Achievement Spring 2014 Scaled Scores, North Carolina: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – Spring 2011 EOG Reading Scaled Scores (*N* = 1,410)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Middle School Cohort									
All	522	347.7	20.10	888	348.1	19.36	-0.37	-0.02	49
Not IEP	467	351.5	10.70	798	351.6	9.19	-0.29	-0.01	50
IEP	55	315.3	41.47	90	316.4	43.05	-0.15	-0.03	49
Not ELL	469	349.7	17.00	825	349.6	17.26	0.09	0.01	50
ELL	53	329.8	32.88	63	328.1	31.10	0.28	0.05	52
Not FRL	196	353.0	15.08	429	353.4	13.48	-0.33	-0.03	49
FRL	326	344.5	21.99	459	343.1	22.47	0.85	0.06	53
Male	260	346.6	20.96	447	345.9	23.36	0.41	0.03	51
Female	262	348.7	19.19	441	350.3	13.90	-1.26	-0.10	46

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Table 47. EOG Mathematics Achievement Spring 2014 Scaled Scores, North Carolina: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – Spring 2011 EOG Mathematics Scaled Scores (*N* = 1,410)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Middle School Cohort									
All	522	351.9	31.64	888	352.5	29.92	-0.33	-0.03	49
Not IEP	467	356.6	15.33	798	356.9	11.07	-0.38	-0.03	49
IEP	55	312.1	76.33	90	313.3	78.12	-0.09	-0.02	49
Not ELL	469	354.0	26.99	825	353.1	29.47	0.54	0.04	52
ELL	53	333.7	55.65	63	344.6	34.59	-1.28	-0.42	34
Not FRL	196	355.9	29.63	429	357.3	22.93	-0.63	-0.07	47
FRL	326	349.5	32.59	459	348.0	34.65	0.63	0.06	52
Male	260	351.8	33.03	447	350.4	37.45	0.50	0.05	52
Female	262	352.0	30.26	441	354.5	19.36	-1.35	-0.16	44

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Table 48. EOG Science Achievement Spring 2014 Scaled Scores, North Carolina: Baseline Subgroup Mean Comparison of Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – Spring 2011 EOG Science Scaled Scores (*N* = 1,409)

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Middle School Cohort									
All	522	157.3	7.45	887	156.9	8.10	0.93	0.05	52
Not IEP	467	157.7	7.29	798	157.2	7.98	1.03	0.06	53
IEP	55	153.6	7.92	89	153.5	8.46	0.05	0.01	50
Not ELL	469	157.7	7.45	824	157.2	8.03	1.11	0.06	53
ELL	53	153	6.10	63	151.8	7.38	0.93	0.17	57
Not FRL	196	159.9	7.55	429	159.7	6.99	0.32	0.03	51
FRL	326	155.7	6.93	458	154.2	8.16	2.69	0.20	58
Male	260	158.2	7.11	446	157.9	8.28	0.56	0.04	52
Female	262	156.3	7.67	441	155.8	7.80	0.80	0.06	53

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Elementary Cohort Reading EOG Spring 2014 Results

For the 1,847 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2011 PASS-Basic scaled scores (Block 3) explained 36% of the total variance (R^2) in students' 2013-2014 reading EOG scaled scores (see Table 49). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 reading scaled scores ($\beta = 0.02$, $t = 0.86$, $p = 0.391$).

The overall ANCOVA analysis (see Table 50) revealed that there was a neither statistically significant nor substantively important difference in students' 2013-2014 reading EOG scaled scores between Phase 1 and Phase 2 elementary cohort students overall. Consistent with the overall outcome, all subgroup ANCOVA analyses revealed neither statistically significant nor substantively important differences.

Table 49. EOG Reading, North Carolina, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 Scaled Scores (N = 1,847)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1842) = 182.84$, $p < 0.001$, $R^2 = 0.268$, F Change $(4,1842) = 182.84$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-30.81	1.26	-0.48	-24.46	< 0.001***
ELL (0 = No, 1 = ELL)	-6.71	1.36	-0.10	-4.95	< 0.001***
FRL (0 = No, 1 = FRL)	-5.95	0.77	-0.16	-7.72	< 0.001***
Gender (0 = M, 1= F)	0.68	0.74	0.02	0.92	0.360
Block 2: Demographics + Fall Score Model Fit: $F(5, 1841) = 202.22$, $p < 0.001$, $R^2 = 0.355$, F Change $(1, 1841) = 200.54$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-27.21	1.22	-0.43	-22.25	< 0.001***
ELL (0 = No, 1 = ELL)	-4.38	1.30	-0.07	-3.37	0.001**
FRL (0 = No, 1 = FRL)	-3.37	0.75	-0.09	-4.47	< 0.001***
Gender (0 = M, 1= F)	1.00	0.70	0.03	1.42	0.155
Fall 2011 Test Score Scaled	0.06	0.00	0.28	14.16	< 0.001***
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 1840) = 168.62$, $p < 0.001$, $R^2 = 0.355$, F Change $(1,1840) = 0.74$, $p = 0.391$					
IEP (0 = No, 1 = IEP)	-27.23	1.22	-0.43	-22.26	< 0.001***
ELL (0 = No, 1 = ELL)	-4.33	1.30	-0.07	-3.32	0.001**
FRL (0 = No, 1 = FRL)	-3.36	0.75	-0.09	-4.46	< 0.001***
Gender (0 = M, 1= F)	0.98	0.70	0.03	1.40	0.161
Fall 2011 Test Score Scaled	0.06	0.00	0.29	14.16	< 0.001***
Phase (0 = P2, 1 = P1)	0.60	0.70	0.02	0.86	0.391

** $p < 0.01$; *** $p < 0.001$

Table 50. EOG Reading, North Carolina, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 1,847)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
North Carolina: Elementary Cohort												
All	886	449.1	17.80	449.2	961	448.7	19.38	448.6	0.74	0.39	0.04	52
Not IEP	795	451.6	11.00	451.6	877	452.0	10.02	452.0	0.97	0.33	-0.04	48
IEP	91	426.6	38.47	424.7	84	414.4	44.44	416.4	1.93	0.17	0.20	58
Not ELL	821	449.7	17.31	450.0	864	449.6	18.39	449.4	0.68	0.41	0.03	51
ELL	65	440.7	21.52	440.9	97	440.6	25.30	440.4	0.03	0.87	0.02	51
Not FRL	478	452.4	15.71	453.0	501	452.5	15.08	452.0	1.56	0.21	0.07	53
FRL	408	445.1	19.26	444.7	460	444.6	22.47	444.9	0.03	0.87	-0.01	50
Male	446	448.8	18.83	449.0	503	447.2	22.08	447.1	3.00	0.08	0.09	54
Female	440	449.4	16.70	449.4	458	450.3	15.76	450.2	0.93	0.34	-0.05	48

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Elementary Cohort Mathematics EOG Spring 2014 Results

For the 1,846 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2011 PASS-Basic scaled scores (Block 3) explained 24% of the total variance (R^2) in students' 2013-2014 mathematics EOG scaled scores (see Table 51). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 mathematics scaled scores (= 0.02, t = 0.87, p = 0.387).

The overall ANCOVA analysis (see Table 52) revealed that there was no statistically significant difference between Phase 1 and Phase 2 elementary cohort students' 2013-2014 mathematics EOG scaled scores overall, and the effect size (g = 0.05) favoring Phase 1 students was not substantively important according to WWC guidelines. Consistent with the overall outcome, all subgroup ANCOVA analyses revealed neither statistically significant nor substantively important differences.

Table 51. EOG Mathematics, North Carolina, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 Scaled Scores (*N* = 1,846)

Source	*B*	*S.E.B.*	β	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1841) = 123.18$, $p < 0.001$, $R^2 = 0.211$, F Change $(4,1841) = 123.18$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-46.75	2.26	-0.43	-20.70	< 0.001***
ELL (0 = No, 1 = ELL)	-8.09	2.43	-0.07	-3.33	0.001**
FRL (0 = No, 1 = FRL)	-7.17	1.38	-0.11	-5.21	< 0.001***
Gender (0 = M, 1= F)	1.19	1.32	0.02	0.90	0.367
Block 2: Demographics + Fall Score Model Fit: $F(5, 1840) = 114.64$, $p < 0.001$, $R^2 = 0.238$, F Change $(1, 1840) = 63.71$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-43.00	2.27	-0.40	-18.94	< 0.001***
ELL (0 = No, 1 = ELL)	-5.65	2.41	-0.05	-2.35	0.019*
FRL (0 = No, 1 = FRL)	-4.48	1.40	-0.07	-3.21	0.001**
Gender (0 = M, 1= F)	1.52	1.30	0.02	1.17	0.240
Fall 2011 Test Score Scaled	0.06	0.01	0.17	7.98	< 0.001***
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 1839) = 95.64$, $p < 0.001$, $R^2 = 0.238$, F Change $(1,1839) = 0.75$, $p = 0.387$					
IEP (0 = No, 1 = IEP)	-43.05	2.27	-0.40	-18.96	< 0.001***
ELL (0 = No, 1 = ELL)	-5.56	2.41	-0.05	-2.31	0.021*
FRL (0 = No, 1 = FRL)	-4.48	1.40	-0.07	-3.21	0.001**
Gender (0 = M, 1= F)	1.50	1.30	0.02	1.15	0.249
Fall 2011 Test Score Scaled	0.06	0.01	0.17	7.99	< 0.001***
Phase (0 = P2, 1 = P1)	1.12	1.30	0.02	0.87	0.387

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 52. EOG Mathematics, North Carolina, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (*N* = 1,846)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
North Carolina: Elementary Cohort												
All	886	446.5	30.80	446.8	960	445.9	32.60	445.7	0.75	0.39	0.05	52
Not IEP	795	450.6	16.30	450.5	877	450.7	14.68	450.7	0.05	0.82	-0.02	49
IEP	91	411.1	74.61	407.9	83	395.2	85.33	398.7	0.61	0.44	0.16	56
Not ELL	821	447.3	29.72	447.6	863	447.0	30.58	446.7	0.56	0.46	0.04	52
ELL	65	437.4	41.29	437.8	97	435.8	45.90	435.6	0.15	0.70	0.06	52
Not FRL	478	450.4	25.48	451.2	501	450.7	24.96	450.0	0.65	0.42	0.07	53
FRL	408	442.0	35.53	441.5	459	440.6	38.63	441.0	0.04	0.84	0.02	51
Male	446	445.6	33.25	446.0	503	443.8	37.37	443.5	1.57	0.21	0.09	54
Female	440	448.1	28.10	447.6	457	448.1	26.24	448.0	0.07	0.79	-0.02	49

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Elementary Cohort Science EOG Spring 2014 Results

For the 1,847 elementary cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2011 PASS-Basic scaled scores (Block 3) explained 10% of the total variance (R^2) in students' 2013-2014 science EOG scaled scores (see Table 53). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 science scaled scores (β = -0.01, t = -0.83, p = 0.526).

The overall ANCOVA analysis (see Table 54) revealed that there was a neither statistically significant nor substantively important difference in students' 2013-2014 science EOG scaled scores between Phase 1 and Phase 2 elementary cohort students overall. Consistent with the overall outcome, all subgroup ANCOVA analyses revealed neither statistically significant nor substantively important differences.

Table 53. EOG Science, North Carolina, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students' 2013-2014 Scaled Scores (N = 1,847)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1842) = 32.16$, $p < 0.001$, $R^2 = 0.065$, F Change $(4,1842) = 32.16$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	13.10	1.30	0.23	10.06	< 0.001***
ELL (0 = No, 1 = ELL)	-1.11	1.40	-0.02	-0.79	0.427
FRL (0 = No, 1 = FRL)	-2.54	0.80	-0.08	-3.19	0.001**
Gender (0 = M, 1= F)	-2.29	0.76	-0.07	-3.00	0.003**
Block 2: Demographics + Fall Score Model Fit: $F(5, 1841) = 39.48$, $p < 0.001$, $R^2 = 0.097$, F Change $(1, 1841) = 64.34$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	15.28	1.31	0.27	11.68	< 0.001***
ELL (0 = No, 1 = ELL)	0.30	1.39	0.01	0.22	0.828
FRL (0 = No, 1 = FRL)	-0.97	0.81	-0.03	-1.21	0.227
Gender (0 = M, 1= F)	-2.09	0.75	-0.06	-2.79	0.005**
Fall 2011 Test Score Scaled	0.03	0.00	0.19	8.02	< 0.001***
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 1840) = 32.96$, $p < 0.001$, $R^2 = 0.097$, F Change $(1,1840) = 0.40$, $p = 0.526$					
IEP (0 = No, 1 = IEP)	15.30	1.31	0.27	11.69	< 0.001***
ELL (0 = No, 1 = ELL)	0.26	1.39	0.00	0.19	0.851
FRL (0 = No, 1 = FRL)	-0.98	0.81	-0.03	-1.21	0.225
Gender (0 = M, 1= F)	-2.08	0.75	-0.06	-2.78	0.006**
Fall 2011 Test Score Scaled	0.03	0.00	0.19	8.02	< 0.001***
Phase (0 = P2, 1 = P1)	-0.47	0.75	-0.01	-0.63	0.526

** $p < 0.01$; *** $p < 0.001$

Table 54. EOG Science, North Carolina, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 1, 847)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
North Carolina: Elementary Cohort												
All	886	255.2	16.32	255.1	961	255.4	17.33	255.5	0.40	0.526	-0.02	49
Not IEP	795	254.1	11.05	254.1	877	254.0	10.16	254.0	0.16	0.687	0.01	51
IEP	91	264.2	38.09	265.5	84	270.7	46.11	269.4	0.36	0.549	-0.15	44
Not ELL	821	255.4	15.71	255.3	864	255.7	16.48	255.8	0.38	0.537	-0.02	49
ELL	65	252.6	22.66	252.4	97	252.8	23.55	253.0	0.03	0.858	-0.01	50
Not FRL	478	256.1	11.72	256.1	501	256.9	13.68	256.9	1.16	0.282	-0.07	47
FRL	408	254.1	20.40	254.2	460	253.8	20.49	253.8	0.12	0.734	0.01	51
Male	446	256.6	17.55	256.3	503	257.0	19.54	257.2	0.60	0.438	-0.02	49
Female	440	253.8	14.86	253.8	458	253.8	14.36	253.8	0.00	0.983	0.00	50

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Middle School Cohort Reading EOG Spring 2014 Results

For the 1,410 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 reading EOG scaled scores (Block 3) explained 70% of the total variance (R^2) in students' 2013-2014 reading EOG scaled scores (see Table 55). The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of 2013-2014 reading scaled scores (β = -0.01, t = -0.40, p = 0.689).

The overall ANCOVA analysis (see Table 56) revealed that there was no statistically significant difference between Phase 1 and Phase 2 middle school cohort students' 2013-2014 reading EOG scaled scores overall, and the effect size (g = -0.02) favoring Phase 2 students was not substantively important according to WWC guidelines. Consistent with the overall outcome, all subgroup ANCOVA analyses revealed neither statistically significant nor substantively important differences.

Table 55. EOG Reading, North Carolina, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 Scaled Scores (N = 1,410)

Source	B	$S.E.B.$	β	t	p
Block 1: Demographics Model Fit: $F(4, 1405) = 181.11$, $p < 0.001$, $R^2 = 0.340$, F Change $(4,1405) = 181.11$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-36.82	1.60	-0.51	-23.02	< 0.001***
ELL (0 = No, 1 = ELL)	-6.85	1.79	-0.09	-3.82	< 0.001***
FRL (0 = No, 1 = FRL)	-7.65	0.99	-0.17	-7.76	< 0.001***
Gender (0 = M, 1= F)	2.04	0.96	0.05	2.12	0.034*
Block 2: Demographics + Fall Score Model Fit: $F(5, 1404) = 654.81$, $p < 0.001$, $R^2 = 0.700$, F Change $(1, 1404) = 1682.53$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-9.10	1.27	-0.13	-7.14	< 0.001***
ELL (0 = No, 1 = ELL)	4.27	1.24	0.05	3.45	0.001**
FRL (0 = No, 1 = FRL)	-2.53	0.68	-0.06	-3.74	< 0.001***
Gender (0 = M, 1= F)	1.33	0.65	0.03	2.05	0.040*
2010 – 2011 reading EOG Scaled Score	0.85	0.02	0.76	41.02	< 0.001***
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 1403) = 545.37$, $p < 0.001$, $R^2 = 0.700$, F Change $(1,1403) = 0.16$, $p = 0.689$					
IEP (0 = No, 1 = IEP)	-9.09	1.27	-0.13	-7.14	< 0.001***
ELL (0 = No, 1 = ELL)	4.29	1.24	0.05	3.46	0.001**
FRL (0 = No, 1 = FRL)	-2.51	0.68	-0.06	-3.69	< 0.001***
Gender (0 = M, 1= F)	1.33	0.65	0.03	2.05	0.040*
2010 – 2011 reading EOG Scaled Score	0.85	0.02	0.76	41.00	< 0.001***
Phase (0 = P2, 1 = P1)	-0.27	0.67	-0.01	-0.40	0.689

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Table 56. EOG Reading, North Carolina, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 1,410)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	p	g	PR
North Carolina: Middle School Cohort												
All	522	455.7	21.50	456.0	888	456.5	22.31	456.3	0.16	0.16	-0.02	49
Not IEP	467	459.6	12.11	459.8	798	460.6	10.16	460.5	3.67	0.06	-0.09	46
IEP	55	422.6	44.13	422.3	90	419.8	50.24	420.0	0.18	0.18	0.06	52
Not ELL	469	457.3	19.53	457.3	825	457.6	21.25	457.6	0.20	0.20	-0.02	49
ELL	53	441.8	31.32	441.2	63	441.4	29.52	441.9	0.03	0.86	-0.03	49
Not FRL	196	462.0	15.54	462.2	429	462.3	14.71	462.2	0.00	0.97	0.00	50
FRL	326	451.9	23.62	451.1	459	451.1	26.48	451.7	0.38	0.54	-0.03	49
Male	260	454.0	22.81	453.5	447	453.7	26.41	454.0	0.19	0.66	-0.02	49
Female	262	457.4	20.02	458.6	441	459.4	16.73	458.7	0.04	0.83	-0.01	50

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (g). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.

Middle School Cohort Mathematics EOG Spring 2014 Results

For the 1,410 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 mathematics EOG scaled scores (Block 3) explained 63% of the total variance (R^2) in students' 2013-2014 mathematics EOG scaled scores (see Table 57). The addition of the student's Phase to the model did not increase the percentage of variance accounted for, and Phase was a statistically significant predictor of 2013-2014 mathematics scaled scores (β= 0.04, t = 2.53, p = 0.011).

The overall ANCOVA analysis (see Table 58) revealed that there was a statistically significant difference in students' 2013-2014 mathematics EOG scaled scores between Phase 1 and Phase 2 middle school cohort students in the aggregate (F (1, 1403) = 6.42, p = 0.011, g = 0.10, *PR* = 54), but the effect size (g = 0.10) favoring Phase 1 students was not substantively important according to WWC guidelines.

The ANCOVA analyses for the subgroup comparisons revealed that Phase 1 students statistically significantly outperformed their Phase 2 counterparts in the Not IEP, Not ELL, Not FRL, and Female subgroups. However, none of the effect sizes associated with these comparisons was substantively important. All other subgroup comparisons were neither statistically significant nor substantively important, with effect size ranging from 0.04 (Male) to 0.20 (IEP). It should be noted that for the ELL subgroup, Phase 2 students had substantively higher baseline scores (g = -0.42). Therefore, it appears that Phase 1 ELL students were able to not only greatly reduce, but even to reverse the achievement gap present at the baseline.

Table 57. EOG Mathematics, North Carolina, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 Scaled Scores (N = 1,410)

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1405) = 98.09$, $p < 0.001$, $R^2 = 0.218$, F Change $(4,1405) = 98.09$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-45.52	2.54	-0.43	-17.94	< 0.001***
ELL (0 = No, 1 = ELL)	-1.17	2.84	-0.01	-0.41	0.679
FRL (0 = No, 1 = FRL)	-8.12	1.56	-0.13	-5.20	< 0.001***
Gender (0 = M, 1= F)	1.78	1.52	0.03	1.17	0.243
Block 2: Demographics + Fall Score Model Fit: $F(5, 1404) = 473.00$, $p < 0.001$, $R^2 = 0.627$, F Change $(1, 1404) = 1542.24$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-13.60	1.93	-0.13	-7.04	< 0.001***
ELL (0 = No, 1 = ELL)	2.61	1.96	0.02	1.33	0.185
FRL (0 = No, 1 = FRL)	-4.31	1.08	-0.07	-3.98	< 0.001***
Gender (0 = M, 1= F)	2.15	1.05	0.03	2.05	0.041*
2010 – 2011 mathematics EOG Scaled Score	0.75	0.02	0.72	39.27	< 0.001***
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 1403) = 396.76$, $p < 0.001$, $R^2 = 0.629$, F Change $(1,1403) = 6.42$, $p = 0.011$					
IEP (0 = No, 1 = IEP)	-13.58	1.93	-0.13	-7.05	< 0.001***
ELL (0 = No, 1 = ELL)	2.44	1.96	0.02	1.24	0.214
FRL (0 = No, 1 = FRL)	-4.57	1.09	-0.07	-4.21	< 0.001***
Gender (0 = M, 1= F)	2.13	1.05	0.03	2.03	0.042*
2010 – 2011 reading EOG Scaled Score	0.75	0.02	0.72	39.33	< 0.001***
Phase (0 = P2, 1 = P1)	2.75	1.09	0.04	2.53	0.011*

* $p < 0.05$; *** $p < 0.001$

Table 58. EOG Mathematics, North Carolina, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 1,410)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
North Carolina: Middle School Cohort												
All	522	445.4	30.20	445.9	888	443.5	33.11	443.2	6.42	0.011*	0.10	54
Not IEP	467	449.4	15.96	449.6	798	448.9	10.83	448.7	7.01	0.008**	0.10	54
IEP	55	411.2	72.58	410.3	90	396.0	85.63	396.5	1.95	0.164	0.20	58
Not ELL	469	447.2	26.05	447.0	825	444.0	32.97	444.1	6.53	0.011*	0.11	54
ELL	53	429.6	52.45	434.8	63	436.5	34.43	432.1	0.69	0.409	0.11	54
Not FRL	196	451.8	17.31	452.3	429	449.7	21.50	449.5	4.73	0.030*	0.16	56
FRL	326	441.5	35.25	440.5	459	437.7	40.26	438.4	1.80	0.180	0.07	53
Male	260	443.2	35.88	442.4	447	440.8	40.69	441.2	0.49	0.483	0.04	51
Female	262	447.5	23.11	449.0	441	446.3	22.73	445.4	7.42	0.007**	0.16	56

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (*g*). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$

Middle School Cohort Science EOG Spring 2014 Results

For the 1,409 middle school cohort students, the hierarchical multiple regression that controlled for student's demographic characteristics and their 2010-2011 science EOG scaled scores (Block 3) explained 25% of the total variance (R^2) in students' 2013-2014 science EOG scaled scores (see Table 59). The addition of the student's Phase to the model (Block 3) increased the percentage of variance accounted for by one percentage point, and Phase was a statistically significant predictor of 2013-2014 science scaled scores (β= -0.05, t = -2.26, p = 0.024).

The overall ANCOVA analysis (see Table 60) revealed that there was a statistically significant difference in students' 2013-2014 science EOG scaled scores between Phase 1 and Phase 2 middle school cohort students in the aggregate (F (1, 1402) = 5.11, p = 0.024, g = -0.09, PR = 46), but the effect size (g = -0.09) favoring Phase 2 students was not substantively important according to WWC guidelines.

The ANCOVA analyses for the subgroup comparisons revealed that Phase 2 students statistically significantly outperformed their Phase 1 counterparts in the Not IEP, Not ELL, and Not FRL subgroups. However, none of the effect sizes associated with these comparisons was substantively important, ranging from -0.21 (Not FRL) to Not ELL (-0.13). In addition, although not statistically significant, the effect size associated with the ELL (g = 0.26) subgroup comparison was substantively important, with the average Phase 1 ELL student scoring at the 60^{th} percentile of the Phase 2 IEP group (PR = 60). All other subgroup comparisons were neither statistically significant nor substantively important, with effect size ranging from -0.12 (Male) to -0.03 (FRL).

Table 59. EOG Science, North Carolina, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students' 2013-2014 Scaled Scores (N = 1,409)

Source	B	S.E.B.	β	t	p
Block 1: Demographics Model Fit: $F(4, 1404) = 67.33$, $p < 0.001$, $R^2 = 0.161$, F Change $(4,1404) = 67.33$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	24.71	1.59	0.39	15.56	< 0.001***
ELL (0 = No, 1 = ELL)	-7.17	1.77	-0.10	-4.04	< 0.001***
FRL (0 = No, 1 = FRL)	-2.89	0.98	-0.07	-2.96	0.003**
Gender (0 = M, 1= F)	-2.44	0.95	-0.06	-2.56	0.010*
Block 2: Demographics + Fall Score Model Fit: $F(5, 1403) = 90.19$, $p < 0.001$, $R^2 = 0.243$, F Change $(1, 1403) = 152.59$, $p < 0.001$					
IEP (0 = No, 1 = IEP)	27.12	1.52	0.43	17.82	< 0.001***
ELL (0 = No, 1 = ELL)	-4.91	1.70	-0.07	-2.89	0.004**
FRL (0 = No, 1 = FRL)	0.54	0.97	0.01	0.56	0.578
Gender (0 = M, 1= F)	-0.62	0.91	-0.02	-0.68	0.499
2010 – 2011 science EOG Scaled Score	0.76	0.06	0.31	12.35	< 0.001***
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 1402) = 76.23$, $p < 0.001$, $R^2 = 0.246$, F Change $(1,1402) = 5.11$, $p = 0.024$					
IEP (0 = No, 1 = IEP)	27.13	1.52	0.43	17.86	< 0.001***
ELL (0 = No, 1 = ELL)	-4.75	1.69	-0.07	-2.80	0.005**
FRL (0 = No, 1 = FRL)	0.78	0.97	0.02	0.81	0.421
Gender (0 = M, 1= F)	-0.58	0.91	-0.02	-0.64	0.524
2010 – 2011 science EOG Scaled Score	0.77	0.06	0.31	12.50	< 0.001***
Phase (0 = P2, 1 = P1)	-2.11	0.94	-0.05	-2.26	0.024*

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 60. EOG Science, North Carolina, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) – 2013-2014 Scaled Scores (N = 1,409)

	Treatment (Phase 1)				Control (Phase 2)							
Group	n	M	SD	Adj. M	n	M	SD	Adj. M	F	p	g	PR
North Carolina: Middle School Cohort												
All	522	252.7	19.84	252.40	887	254.4	18.99	254.52	5.11	0.024*	-0.09	46
Not IEP	467	250.2	10.31	250.15	798	252.0	9.88	251.99	16.70	< 0.001***	-0.18	43
IEP	55	273.5	48.82	273.86	89	276.0	47.12	275.78	0.06	0.812	-0.05	48
Not ELL	469	252.5	17.96	252.24	824	255.0	18.84	255.11	9.29	0.002**	-0.13	45
ELL	53	253.6	32.20	253.04	63	246.9	19.45	247.37	1.81	0.181	0.26	60
Not FRL	196	253.2	13.13	253.01	429	255.8	12.68	255.87	7.82	0.005**	-0.21	42
FRL	326	252.3	22.96	252.05	458	253.0	23.34	253.24	0.71	0.399	-0.03	49
Male	260	254.1	20.60	254.37	446	256.6	21.96	256.44	1.90	0.169	-0.12	45
Female	262	251.2	18.98	250.45	441	252.1	15.10	252.58	3.57	0.059	-0.05	48

Note: PR = The percentile rank of the average Phase 1 student in the control group based on the effect size (g). For example, if the PR is 60, then the average Phase 1 student scored at the 60th percentile of the control group.
* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

References

What Works Clearinghouse (2014). Procedures and standards handbook (Version 3.0). Washington, DC: Author. Retrieved from ies.ed.gov/ncee/wwc/pdf/reference_resources/ wwc_procedures_v3_0_standards_handbook.pdf