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 4:

PASS Assessments Open Ended and Performance Task

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 – toimprove 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

A random sample of schools in the three regions took the PASS Open-Ended and Performance Task assessment for the first time in Spring 2012, (end of first posttest year) and again in Spring 2013 and Spring 2014 (second and third posttest years, respectively). Students in the elementary cohort (currently 5th graders in 2013-2014) responded to two Open-Ended (OE) and six Performance Task (PT) items, while students in the middle school cohort (currently 8th graders in 2013-2014) responded to six OE and six PT items. It should be noted that a random sample of schools in the HISD middle school cohort took the OE and PT sections for the first time in Spring 2013, and are therefore not included in these analyses.

PASS Open Ended and Performance Task Scoring

For the elementary cohort, there are a total of six points possible for the OE section and 17 total points possible for the PT section. For the middle school cohort, there are a total of 15 points possible for the OE section and 17 total points possible for the PT section. The items are scored using a rubric, with the number of points available for each item in each section shown in Table 1 below. In order to score a section, the student had to answer at least one item (i.e., gave a response that received a score of zero or higher). Otherwise, the section was dropped from the analysis if all the items were either missing, scored a "B" (blank), or had a combination of missing data and scores of "B". If the section was scored, any item with a "B" and any missing items were given a value of zero. As a result, when a section was scored and a student had missing items or items scored with a "B", those items were treated the same as the case where a student actually responded to an item, but received a score of zero, indicating the response did not contain any correct elements or was irrelevant. For both the OE and PT sections, the outcome score used in the analyses was the percentage correct out of the total number of points possible.

Table 1. PASS OE and PT Scoring Scales, Spring 2012, Spring 2013, and Spring 2014

Elementary Cohort				Middle School Cohort			
Open-ended Question		Performance Task		Open-ended Question		Performance Task	
Item	Scale	Item	Scale	Item	Scale	Item	Scale
1	B, 0, 1, 2, 3	1	B, 0, 1, 2, 3	1	B, 0, 1, 2	1	B, 0, 1, 2, 3
2	B, 0, 1, 2, 3	2	B, 0, 1, 2, 3	2	B, 0, 1, 2	2	B, 0, 1, 2, 3
		3	B, 0, 1, 2, 3	3	B, 0, 1, 2	3	B, 0, 1, 2, 3
		4	B, 0, 1, 2, 3	4	B, 0, 1, 2, 3	4	B, 0, 1, 2, 3
		5	B, 0, 1, 2, 3	5	B, 0, 1, 2, 3	5	B, 0, 1, 2, 3
		6	B, 0, 1, 2	6	B, 0, 1, 2, 3	6	B, 0, 1, 2
Total Points	6	Total Points	17	Total Points	15	Total Points	17

B = Blank

Results for All Regions combined are presented first, followed by the outcomes for the Houston Independent School District (HISD), the New Mexico region, and the North Carolina region. 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), outcome (PASS OE and PASS PT) and subgroup.

All Regions:
Results for Spring 2014 PASS
Open-Ended and Performance Task

All Regions Spring 2014 PASS Open-Ended and Performance Task Key Findings for Phase 1

For all students combined (the "All" group) and the specified subgroups, the following outcomes favoring Phase 1 students were found on the Spring 2014 PASS Open-Ended and Constructed Response sections.

ELL

- Elementary Cohort Performance Task: After controlling for the statistically significant advantage Phase 2 students demonstrated on the pretest (g = -0.18), Phase 1 students demonstrated a statistically significant and substantively important advantage on the posttest over Phase 2 students (g = 0.30).
- Middle School Cohort Performance Task: Phase 1 students had a substantially important advantage over Phase 2 students on the posttest (g = 0.37).

Economically Disadvantaged (FRL)

- Middle School Cohort Performance Task: Phase 1 students statistically significantly outperformed Phase 2 students with an effect size that was substantively important (g = 0.27).

IEP

- Elementary Cohort Performance Task: Phase 1 students statistically significantly outperformed Phase 2 students on the posttest with a substantially important effect size (g = 0.39).

Female

- Middle School Cohort Performance Task: Phase 1 students statistically significantly outperformed Phase 2 students with a nearly substantively important effect size (g = 0.23).

Spring 2014 PASS Open-Ended and Performance Task Results: All Regions

A preliminary analysis was conducted on the Spring 2012 OE and PT sections of the PASS to determine baseline equivalence between Phase 1 and Phase 2 students in the elementary and middle school cohorts included the present analysis (see Table 2) as the PASS OE and PT sections were not administered until the end of the first posttest year, meaning there was no Fall 2011 baseline scores available. 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, 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).

Results indicated that for the elementary cohort aggregate scores (i.e., for all students combined), there was no statistically significant difference between Phase 1 and Phase 2 on the Spring 2012 OE or PT percent correct, along with no substantially important effect sizes according to What Work Clearinghouse (WWC) standards. For the middle school cohort aggregate scores, Phase 1 students had a statistically significantly higher mean Spring 2012 OE percent correct, as well as Spring 2012 PT percent correct, with the magnitude of the effects for both being substantially important.

Table 2. PASS OE and PT, Spring 2012 For Students Who Had a Spring 14 OE or PT Score, Treatment (Phase 1) and Control (Phase 2) Means Comparison: All Regions

		Treatment (Phase 1)			Control (Phase 2)				
Section	Cohort	n	M	SD	n	M	SD	t	*g*
Open-Ended	Elementary	1159	43.3	20.66	991	44.43	18.69	-1.37	-0.06
Performance Task	Elementary	1326	53.68	19.76	1099	54.41	17.35	-0.97	-0.04
Open-Ended	Middle School	795	72.6	16.32	578	68.06	19.43	4.56*	0.26
Performance Task	Middle School	697	52.11	20.08	514	42.23	23.45	7.69*	0.46

* $p < 0.05$

Due to the fact that the PASS OE and PT were not administered until the end of the first posttest year, meaning there were no true baseline scores available, and due to substantively meaningful differences on the Spring 2012 scores, correlation analyses were conducted to examine the relationship between the Spring 2014 PASS OE and PT percent correct and (1) the Spring 2012 PASS OE and PT percent correct, as well as (2) the Fall 2011 PASS Multiple Choice (MC) scaled score results, to determine which scores would serve as the better baseline measure of achievement. The analyses revealed statistically significant, but low correlations among each of the measures of achievement (see Table 3). For both the elementary and middle school cohorts, the Fall 2011 PASS MC scaled scores had higher statistically significant correlations with the Spring 2014 PASS OE and PT, compared to the Spring 2012 OE and PT.

Table 3. Correlations on the Percent Correct for Spring 2014 PASS OE and PT with Spring 2012 PASS OE and PT, and Fall 2011 PASS Multiple Choice: All Regions

Spring 2014 PASS	Cohort	Fall 2011 PASS Multiple Choice	Spring 2012 Open-Ended	Spring 2012 Performance Task
Spring 2014 Open-Ended	Elementary	0.37*	0.33*	NA
	Middle School	0.45*	0.38*	NA
Spring 2014 Performance Task	Elementary	0.36*	NA	0.35*
	Middle School	0.39*	NA	0.34*

* $p < 0.05$

To determine baseline equivalence on the Fall 2011 PASS MC between Phase 1 and Phase 2 students included the present analyses, a series of independent *t*-tests were 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 (see Table 4). For the elementary OE cohort in the aggregate (i.e., the "All" group), Phase 2 students demonstrated a statistically significant advantage over their Phase 1 counterparts in their baseline achievement levels ($t(2583) = -2.53$, $p = 0.011$, $g = -0.10$, PR = 46), but the effect size linked to this advantage did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$). Consistent with this overall difference in performance, statistically significant, but not substantively important advantages were observed to favor two subgroups of Phase 2 students in the elementary cohort: Not IEP and Females.

For students in the middle school OE cohort, no statistically significant difference in aggregate performance (i.e., the "All" group) between Phase 1 and Phase 2 students was observed ($t(1525) = 1.02$, $p = 0.309$, $g = 0.05$, PR = 52), and was linked to an effect size that did not meet the WWC criteria for substantive importance. Meanwhile, a statistically significant, but not substantively important advantage in baseline performance was observed for the subgroup of Phase 1 middle school cohort students who were considered FRL. Additionally, ELL Phase 1 students had an advantage over Phase 2 students that was not statistically significant, but was substantively important ($g = 0.31$).

With respect to the elementary PT cohort in the aggregate (i.e., the "All" group), Phase 2 students demonstrated a statistically significant advantage over their Phase 1 counterparts in their baseline achievement levels ($t(2599) = -2.20$, $p = 0.028$, $g = -0.09$, PR = 47), but the effect size linked to this advantage did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$). Consistent with this overall difference in performance, statistically significant, but not substantively important advantages were observed to favor two subgroups of Phase 2 students in the elementary cohort: Not IEP and ELL.

With respect to students in the middle school PT cohort, no statistically significant difference in aggregate performance (i.e., the "All" group) between Phase 1 and Phase 2 students was observed ($t(1406) = -0.55$, $p = 0.582$, $g = -0.03$, PR = 49), and the associated effect size did not meet the WWC criteria for substantive importance. No statistically significant or substantively important advantages in baseline performance were observed for any of the eight subgroups of middle school cohort students.

While neither the Spring 2012 PASS OE and PT nor the Fall 2011 PASS Multiple Choice provided complete baseline equivalence between Phase 1 and Phase 2 students, the Fall 2011 PASS Multiple Choice was administered as a true baseline assessment vs. the Spring 2012 PASS OE and PT, which was not administered until the end of the first posttest year. Therefore, due to its stronger relationship to the Spring 2014 PASS OE and PT outcomes, and because it was a true baseline measure, the Fall 2011 PASS Multiple Choice scaled score was chosen as the covariate (i.e., pretest measure) for both the elementary and middle school cohorts.

Table 4. Fall 2011 PASS Multiple Choice for Students Who Had Spring 2014 OE or PT Scores, Treatment (Phase 1) and Control (Phase 2) Means Comparison: All Regions

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort - Open-Ended									
All	1409	306.03	100.39	1176	316	98.81	-2.53*	-0.10	46
Not IEP	1276	310.23	100.27	1086	320.77	96.99	-2.58*	-0.11	46
IEP	133	265.75	92.62	90	258.42	102.86	0.55	0.08	53
Not ELL	1039	321.12	98.96	929	326.12	99.47	-1.12	-0.05	48
ELL	370	263.68	92.01	247	277.96	86.41	-1.94	-0.16	44
Not FRL	519	350.55	92.38	517	352.31	98.63	-0.30	-0.02	49
FRL	890	280.08	95.69	659	287.52	89.2	-1.56	-0.08	47
Male	716	309.72	105.76	592	318.92	100.23	-1.3	-0.09	46
Female	693	302.23	94.44	584	313.05	97.34	-2.01*	-0.11	46
	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Middle School Cohort - Open-Ended									
All	832	368.45	103.45	695	362.84	111.7	1.02	0.05	52
Not IEP	749	378.75	99.68	624	373.45	105.64	0.95	0.05	52
IEP	83	275.46	90	71	269.59	120.65	0.34	0.06	52
Not ELL	788	373.71	102.09	631	374.62	106.79	-0.16	-0.01	50
ELL	44	274.25	80.46	64	246.67	91.15	1.62	0.31	62
Not FRL	342	410.75	93.95	315	416.65	89.45	-0.82	-0.06	47
FRL	490	338.92	99.53	380	318.23	108.71	2.92*	0.20	58
Male	398	364.5	107.88	343	366.19	115.03	-0.21	-0.02	49
Female	434	372.06	99.2	352	359.58	108.43	1.68	0.12	55
	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort - Performance Task									
All	1429	308.05	101.59	1172	316.73	98.39	-2.20*	-0.09	47
Not IEP	1297	312.27	101.63	1078	322.14	96.4	-2.41*	-0.10	46
IEP	132	266.52	91.6	94	254.68	100.22	0.92	0.12	55
Not ELL	1058	323.66	100.07	934	326.19	99.21	-0.57	-0.03	49
ELL	371	263.52	92.41	238	279.6	85.71	-2.15*	-0.18	43
Not FRL	534	354.65	94.29	518	352.23	99.07	0.41	0.02	51
FRL	895	280.24	95.46	654	288.61	88.3	-1.76	-0.09	46
Male	726	312.09	107.55	591	319.18	100.5	-1.23	-0.07	47
Female	703	303.87	94.93	581	314.23	96.22	-1.93	-0.11	46

Table 4, continued

Group	Treatment (Phase 1) *n*	*M*	*SD*	Control (Phase 2) *n*	*M*	*SD*	*t*	*g*	PR
Middle School Cohort - Performance Task									
All	772	365.64	104.81	636	368.78	107.97	-0.55	-0.03	49
Not IEP	688	377.09	100.63	575	378.1	103.06	-0.18	-0.01	50
IEP	84	271.93	90.84	61	280.89	114.37	-0.53	-0.09	46
Not ELL	730	370.86	103.63	586	378.14	104.13	-1.26	-0.07	47
ELL	42	274.98	82.04	50	259.06	91.01	0.87	0.08	57
Not FRL	307	406.08	98	298	417.36	86.86	-1.5	-0.12	45
FRL	465	338.95	100.62	338	325.94	106.72	1.76	0.13	55
Male	367	362.65	107.75	308	375.42	109.99	-1.52	-0.12	45
Female	405	368.36	102.13	328	362.54	105.82	0.75	0.06	52

* $p < 0.05$

Employing the Fall 2011 PASS MC data as covariates to statistically adjust the outcomes for baseline differences in achievement, preliminary analyses were conducted on Spring 2014 PASS OE and PT percent correct scores to determine any differences between Phase 1 and Phase 2 students in the elementary and middle school cohorts. As these analyses were exploratory in nature, no corrections were made for multiple comparisons. As noted earlier, for the elementary cohort, there were statistically significant differences in Phase 1 and Phase 2 students on the baseline measures for both the OE and PT, with Phase 2 students having an advantage both overall and for several subgroups. For the middle school cohort, Phase 1 ELL students had a substantively important advantage and FRL students had a statistically significant advantage on the OE section. Due to these baseline differences, ANCOVA assumptions for equal variances were violated. Therefore, results for these particular groups should be interpreted with these advantages in mind.

Elementary and Middle School Cohorts PASS Open-Ended Analyses: All Regions

With respect to the 2,585 elementary cohort students in Phase 1 (n = 1,409) and Phase 2 (n = 1,176) schools and the 1,527 middle school cohort students in Phase 1 (n = 832) and Phase 2 (n =695) schools, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their percent correct scores on the Spring 2014 OE section of the PASS assessment. In addition to these regressions, a second set of ANCOVA analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (g), was also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their Special Education (IEP) status, English Language Learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender.

Elementary Cohort Spring 2014 PASS Open-Ended Results: All Regions

Among the 2,585 elementary cohort students across the three regions, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS Multiple Choice scaled scores (Block 3) explained 17% of the total variance (R^2) in students' Spring 2014 OE scores (see Table 5), with the addition of Phase to the model not adding appreciably to the percentage of variance explained. However, Phase did have a statistically significant impact on the Spring 2014 OE percent correct (β = -0.05, t = -2.48, p = 0.013) favoring Phase 1.

For the ANCOVA adjusted means presented in Table 6, Phase 1 students (n = 1,409, Adjusted Mean = 66.39) had statistically significantly higher posttest means compared to Phase 2 students (n = 1,176, Adjusted Mean = 64.50) overall (i.e., the "All" group) (F (1, 2578) = 6.32, p = 0.012, g = 0.09, PR = 54). Although the difference was statistically significant, the magnitude of the effect size (g = 0.09) was not considered to be substantively important. It should be noted that on the pretest, Phase 2 students had a statistically significant advantage over Phase 1 students, although it was not considered substantively important. Consistent with these overall outcomes, three subgroup analyses (Not IEP, ELL, and FRL) were statistically significant and all favored Phase 1 elementary cohort students, but did not have corresponding effect sizes that were substantively important. Even though Phase 2 students had an advantage on the pretest overall and for all but the ELL subgroup, after controlling for pretest differences, Phase 1 students outperformed Phase 2 students on the posttest, although no posttest effect size was substantively important for any subgroup.

Table 5. PASS Open-Ended Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Elementary Cohort (N = 2,585): All Regions

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 2581) = 50.99$, $p < .001$, $R^2 = .07$, F Change $(4, 2581) = 50.99$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-12.74	1.40	-0.17	-907	<0.001*
ELL (0 = No, 1 = ELL)	-4.37	1.00	-0.09	-4.38	<0.001*
FRL (0 = No, 1 = FRL)	-6.22	0.67	-0.15	-7.17	<0.001*
Gender (0 = M, 1= F)	3.27	0.79	0.08	4.16	<0.001*
Block 2: Demographics + Fall 2011 PASS Score Model Fit: $F(5, 2580) = 100.50$, $p < .001$, $R^2 = .16$, F Change $(1, 2580) = 276.74$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-8.62	1.36	-0.12	-6.35	<0.001*
ELL (0 = No, 1 = ELL)	-2.27	0.96	-0.05	-2.38	0.018*
FRL (0 = No, 1 = FRL)	-2.22	0.86	-0.05	-2.58	0.010*
Gender (0 = M, 1= F)	3.81	0.75	0.09	5.10	<0.001*
Fall 2011 PASS Scaled Score	0.07	0.00	0.33	16.64	<0.001*
Block 3: Demographics + Fall 2011 PASS Score + Phase Model Fit: $F(6, 2579) = 84.95$, $p < .001$, $R^2 = .17$, F Change $(1, 2579) = 6.17$, $p = .013$)					
IEP (0 = No, 1 = IEP)	-8.73	1.36	-0.12	-6.44	<0.001*
ELL (0 = No, 1 = ELL)	-2.36	0.96	-0.05	-2.47	0.014*
FRL (0 = No, 1 = FRL)	-2.31	0.86	-0.06	-2.70	0.007*
Gender (0 = M, 1= F)	3.82	0.75	0.09	5.11	<0.001*
Fall 2011 PASS Scaled Score	0.07	0.00	0.33	16.69	<0.001*
Phase (1= P1, 2 = P2)	-1.86	0.75	-0.05	-2.48	0.013*

$p < 0.05$

Table 6. PASS Open-Ended Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Elementary Cohort (N = 2,585): All Regions

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	1,409	65.87	21.09	66.39	1,176	65.12	20.18	64.50	6.32	0.012*	0.09	54
Not IEP	1,276	66.88	20.63	67.36	1,086	66.25	19.60	65.69	4.65	0.031*	0.08	53
IEP	133	56.27	22.99	55.97	90	51.48	22.16	51.92	1.91	0.168	0.18	57
Not ELL	1,039	67.29	20.43	67.63	929	66.81	19.76	66.43	2.04	0.154	0.06	52
ELL	370	61.89	22.38	62.38	247	58.77	20.54	58.05	6.70	0.010*	0.20	58
Not FRL	519	70.07	19.32	70.39	517	69.73	19.01	69.41	0.78	0.378	0.05	52
FRL	890	63.43	21.69	63.69	659	61.51	20.35	61.15	6.34	0.012*	0.12	55
Male	716	64.13	21.90	64.61	592	63.26	20.71	62.68	3.14	0.077	0.09	54
Female	693	67.68	20.07	68.21	584	67.01	19.47	66.37	3.22	0.073	0.09	54

* $p < 0.05$.

Middle School Cohort Spring 2014 PASS Open-Ended Results: All Regions

Among the 1,527 middle school cohort students across the three regions, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS MC scaled scores (Block 3) explained 27% of the total variance (R^2) in students' 2014 Spring OE scores, with the addition of Phase to the model not adding to the percentage of variance explained. In addition, Phase did not have a statistically significant impact on the Spring 2014 OE percent correct, (β = -0.02, t = -0.71, p = 0.477) (see Table 7).

The ANCOVA adjusted means presented in Table 8 indicate no statistically significant difference between Phase 1 students (n = 832, Adjusted Mean = 85.08) and Phase 2 students (n = 695, Adjusted Mean = 84.60) overall (i.e., the "All" group) (F (1, 1520) = 0.51, p = 0.477, g = 0.03, PR = 51). No subgroup comparison was statistically significant, although the ELL subgroup produced an effect size that was substantively important favoring Phase 2 (g = -0.32) (with Phase 1 ELL students having an advantage (g = 0.31) on the pretest). However, after controlling for the advantage of Phase 2 Not ELL (g = -0.01) and Male (g =-0.02) students on the pretest, Phase 1 Not ELL and Male students were able to demonstrate a small, but positive effect size (g = 0.07 and g = 0.02 respectively) on the posttest.

Table 7. PASS Open-Ended Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Middle School Cohort (N = 1,527): All Regions

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1522) = 81.64$, $p < .001$, $R^2 = .18$, F Change $(4, 1522) = 81.64$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-12.25	1.25	-0.24	-9.83	<0.001*
ELL (0 = No, 1 = ELL)	-9.83	1.47	-0.16	-6.68	<0.001*
FRL (0 = No, 1 = FRL)	-5.56	0.74	-0.18	-7.49	<0.001*
Gender (0 = M, 1= F)	3.20	0.72	0.10	4.42	<0.001*
Block 2: Demographics + Fall 2011 PASS MC Score Model Fit: $F(5, 1521) = 112.10$, $p < .001$, $R^2 = .27$, F Change $(1, 1521) = 192.80$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-8.45	1.21	-0.17	-7.01	<0.001*
ELL (0 = No, 1 = ELL)	-6.54	1.41	-0.11	-4.65	<0.001*
FRL (0 = No, 1 = FRL)	-2.02	0.74	-0.07	-2.71	0.007*
Gender (0 = M, 1= F)	3.47	0.68	0.11	5.08	<0.001*
Fall 2011 PASS MC SS	0.05	0.00	0.35	13.89	<0.001*
Block 3: Demographics + Fall 2011 PASS MC Score + Phase Model Fit: $F(6, 1520) = 93.47$, $p < .001$, $R^2 = .27$, F Change $(1, 1520) = 0.51$, $p = .477$)					
IEP (0 = No, 1 = IEP)	-8.47	1.21	-0.17	-7.02	<0.001*
ELL (0 = No, 1 = ELL)	-6.46	1.41	-0.11	-4.58	<0.001*
FRL (0 = No, 1 = FRL)	-2.05	0.75	-0.07	-2.75	0.006*
Gender (0 = M, 1= F)	3.47	0.68	0.11	5.07	<0.001*
Fall 2011 PASS MC SS	0.05	0.01	0.35	13.85	<0.001*
Phase (1= P1, 2 = P2)	-0.49	0.68	-0.02	-0.71	0.477

$p < 0.05$

Table 8. PASS Open-Ended Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Middle School Cohort (N = 1,527): All Regions

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	832	85.32	15.49	85.08	695	84.32	15.32	84.60	0.51	0.477	0.03	51
Not IEP	749	86.92	14.04	86.72	624	86.00	13.09	86.24	0.51	0.475	0.04	51
IEP	83	70.92	20.04	70.28	71	69.48	23.55	70.24	0.00	0.992	0.00	50
Not ELL	788	86.36	14.57	86.45	631	85.53	14.22	85.42	2.30	0.129	0.07	53
ELL	44	66.67	19.40	66.25	64	72.40	20.12	72.68	2.77	0.099	-0.32	37
Not FRL	342	88.77	12.75	88.91	315	89.50	10.98	89.35	0.27	0.602	-0.04	49
FRL	490	82.91	16.74	63.69	380	80.02	17.01	61.15	1.09	0.297	0.15	56
Male	716	64.13	21.90	82.85	592	63.26	20.71	82.39	0.19	0.665	0.02	51
Female	693	67.68	20.07	68.21	584	67.01	19.47	66.37	0.36	0.549	0.09	54

* $p < 0.05$.

Elementary and Middle School Cohorts PASS Performance Task Analyses: All Regions

With respect to the 2,601 elementary cohort students in Phase 1 (n = 1,429) and Phase 2 (n = 1,172) schools and the 1,408 middle school cohort students in Phase 1 (n = 772) and Phase 2 (n = 636) schools, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their percent correct score on the Spring 2014 PT section of the PASS assessment. In addition to these regressions, a second set of ANCOVA analyses intended to generate pairs of adjusted percentage correct scores and to compute the treatment effect sizes (g), was also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their Special Education (IEP) status, English Language Learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender.

Elementary Cohort Spring 2014 PASS Performance Task Results: All Regions

Among the 2,601 elementary cohort students across the three regions, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS Multiple Choice scaled scores (Block 3) explained 17% of the total variance (R^2) in students' Spring 2014 PT scores (see Table 9), with the addition of Phase not adding to the percentage of variance explained. There was, however, a statistically significant difference in the Spring 2014 PT percent correct favoring Phase 1 students (β = -0.04, t = -2.45, p = 0.014).

The ANCOVA adjusted means presented in Table 10 can be seen to be statistically significantly higher for Phase 1 students (n = 1,429, Adjusted Mean = 66.55) than for Phase 2 students (n = 1,172, Adjusted Mean = 65.09) overall (i.e., the "All" group) (F (1, 2594) = 6.28, p = 0.012, g = 0.09, PR = 54), indicating that the average Phase 1 student scored at the 54th percentile of the control group. However, the effect size was not considered to be substantively important according to WWC standards. Consistent with these overall outcomes, two subgroup analyses (FRL and Male) were statistically significant and favored Phase 1 elementary cohort students, but did not have substantively import effects, after controlling for baseline Phase 2 advantages. Additionally, two subgroups (IEP and ELL) had statistically significant differences favoring Phase 1 students, along with effect sizes that were substantively important (IEP, g = 0.39 and ELL, g = 0.30).

Meanwhile, even though Phase 2 students had a statistically significant advantage on the pretest overall and for two subgroups (Not IEP and ELL), for the All group and the ELL subgroup, after controlling for statistically significant pretest differences (All, g = -0.09, and ELL, g = -0.18), Phase 1 students statistically significantly outperformed Phase 2 students on the posttest, with the ELL subgroup having a substantively important posttest effect size (g = 0.30). For the Not IEP subgroup, after controlling for the statistically significant Phase 2 pretest advantage (g = -0.10), Phase 1 students had a posttest advantage, but it was neither not statistically significant nor substantially important (g = 0.06). In addition, after controlling for the advantage of Phase 2 Not ELL (g = -0.03) and Female (g =-0.11) students on the pretest, Phase 1 Not ELL and Female students were able to demonstrate a small, but positive effect size (g = 0.03 and g = 0.06 respectively) on the posttest.

Table 9. PASS Performance Task Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Elementary Cohort (N = 2,601): All Regions

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 2597) = 65.28$, $p < .001$, $R^2 = .09$, F Change $(4, 2597) = 65.28$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-10.16	1.08	-0.18	-9.43	<0.001*
ELL (0 = No, 1 = ELL)	-3.29	0.77	-0.09	-4.25	<0.001*
FRL (0 = No, 1 = FRL)	-6.39	0.67	-0.19	-9.59	<0.001*
Gender (0 = M, 1= F)	2.29	0.61	0.07	3.77	<0.001*
Block 2: Demographics + Fall PASS MC Score Model Fit: $F(5, 2596) = 103.44$, $p < .001$, $R^2 = .17$, F Change $(1, 2596) = 232.79$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-7.13	1.05	-0.12	-678	<0.001*
ELL (0 = No, 1 = ELL)	-1.80	0.75	-0.05	-2.40	0.016*
FRL (0 = No, 1 = FRL)	-3.48	0.67	-0.11	-5.22	<0.001*
Gender (0 = M, 1= F)	2.67	0.58	0.08	4.60	<0.001*
Fall PASS MC Scaled Score	0.05	0.00	0.30	15.26	<0.001*
Block 3: Demographics + Fall PASS MC Score + Phase Model Fit: $F(6, 2595) = 87.67$, $p < .001$, $R^2 = .17$, F Change $(1, 2595) = 6.00$, $p = .014$)					
IEP (0 = No, 1 = IEP)	-7.19	1.05	-0.13	-6.84	<0.001*
ELL (0 = No, 1 = ELL)	-1.88	0.75	-0.05	-2.51	0.012*
FRL (0 = No, 1 = FRL)	-3.55	0.67	-0.11	-5.32	<0.001*
Gender (0 = M, 1= F)	2.68	0.58	0.08	4.61	<0.001*
Fall PASS MC Scaled Score	0.05	0.00	0.30	15.30	<0.001*
Phase (1= P1, 2 = P2)	-1.43	0.58	-0.04	-2.45	0.014*

* $p < 0.05$

Table 10. PASS Performance Task Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Elementary Cohort (N = 2,601): All Regions

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	1,429	66.16	15.50	66.55	1,172	65.56	16.82	65.09	6.28	0.012*	0.09	54
Not IEP	1,297	66.76	15.30	67.13	1,078	66.72	15.88	66.27	2.10	0.147	0.06	52
IEP	132	60.29	16.26	59.96	94	52.25	21.16	52.72	10.16	0.002*	0.39	65
Not ELL	1,058	67.20	15.60	67.40	934	67.19	16.13	66.96	0.45	0.504	0.03	51
ELL	371	63.20	14.85	63.55	238	59.17	17.95	58.63	15.54	<0.001*	0.30	62
Not FRL	534	70.19	14.72	70.24	518	70.15	14.61	70.10	0.03	0.874	0.01	50
FRL	895	63.75	15.47	63.95	654	61.93	17.56	61.66	8.37	0.004*	0.14	56
Male	726	65.20	16.08	65.54	591	63.93	17.21	63.51	5.81	0.016*	0.12	55
Female	703	67.15	14.82	67.57	581	67.22	16.26	66.70	1.18	0.278	0.06	52

* $p < 0.05$.

Middle School Cohort Spring 2014 PASS Performance Task Results: All Regions

Among the 1,408 middle school cohort students across the three regions, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS MC scaled scores (Block 3) explained 19% of the total variance (R^2) in students' 2014 Spring PT scores (see Table 11) with the addition of Phase to the model not adding appreciably to the percentage of variance explained.. Phase also had a statistically significant impact on the Spring 2014 PT percent correct, favoring Phase 1 students (β = -0.11, t = -4.37, p <0.001).

The ANCOVA adjusted means presented in Table 12 can be seen to be statistically significantly higher for Phase 1 students (n = 772, Adjusted Mean = 58.81) than for Phase 2 students (n = 636, Adjusted Mean = 53.74) overall (i.e., the "All" group) (F (1, 1401) = 19.09, p <0.001, g = 0.12, PR = 58), indicating that the average Phase 1 student scored at the 58th percentile of the control group. However, the effect size was not considered to be substantively important according to WWC standards. In addition, for the All group, as well as five subgroups (Not IEP, IEP, Not ELL, Not FRL, Male), after controlling for a Phase 2 advantage on the pretest, Phase 1 students statistically significantly outperformed Phase 2 students on the posttest for the All group, as well as four subgroups (Not IEP, Not ELL, Not FRL, and Male), while the IEP subgroup showed a small but positive posttest effect size (g = 0.19). Meanwhile, although three additional subgroups (ELL, FRL, and Female) had Phase 1 advantages on the pretest that were neither statistically significant nor substantially important (ELL, g = 0.08, FRL, g = 0.13, and Female, g = 0.06) the Phase 1 advantages were even stronger on the posttest with substantially important effects for the ELL subgroup (g = 0.37), statistically significant and substantially important effects for the FRL subgroup (g = 0.27), and statistically significant, and nearly substantially important effects for the Female subgroup (g = 0.23).

Table 11. PASS Performance Task Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Middle School Cohort (N = 1,408): All Regions

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1403) = 38.40$, $p < .001$, $R^2 = .10$, F Change $(4, 1403) = 38.40$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-15.06	2.08	-0.19	-7.24	<0.001*
ELL (0 = No, 1 = ELL)	-9.44	2.57	-0.10	-3.67	<0.001*
FRL (0 = No, 1 = FRL)	-6.96	1.25	-0.14	-5.55	<0.001*
Gender (0 = M, 1= F)	3.83	1.23	0.08	3.13	0.002*
Block 2: Demographics + Fall 2011 PASS MC Score Model Fit: $F(5, 1402) = 61.68$, $p < .001$, $R^2 = .18$, F Change $(1, 1402) = 139.65$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-9.23	2.04	-0.12	-4.52	<0.001*
ELL (0 = No, 1 = ELL)	-5.23	2.48	-0.05	-2.11	0.035*
FRL (0 = No, 1 = FRL)	-2.12	1.26	-0.04	-1.68	0.094
Gender (0 = M, 1= F)	4.54	1.17	0.10	3.88	<0.001*
Fall PASS MC Scaled Score	0.07	0.01	0.32	11.82	<0.001*
Block 3: Demographics + Fall 2011 PASS MC Score + Phase Model Fit: $F(6, 1401) = 55.25$, $p < .001$, $R^2 = .19$, F Change $(1, 1401) = 19.09$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-9.50	2.03	-0.12	-4.68	<0.001*
ELL (0 = No, 1 = ELL)	-4.50	2.47	-0.05	-1.83	0.068
FRL (0 = No, 1 = FRL)	-2.55	1.26	-0.05	-2.02	0.043*
Gender (0 = M, 1= F)	4.49	1.16	0.09	3.86	<0.001*
Fall PASS MC Scaled Score	0.07	0.01	0.32	11.86	<0.001*
Phase (1= P1, 2 = P2)	-5.07	1.16	-0.11	-4.37	<0.001*

$p < 0.05$

Table 12. PASS Performance Task Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Middle School Cohort (N = 1,408): All Regions

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	772	58.64	24.49	58.81	636	53.95	23.01	53.74	19.09	<0.001*	0.12	58
Not IEP	688	60.80	24.08	60.86	575	55.71	22.64	55.64	17.75	<0.001*	0.22	59
IEP	84	40.97	20.43	41.02	61	37.32	19.65	37.25	1.28	0.260	0.19	57
Not ELL	730	59.42	24.48	59.81	586	55.40	22.78	54.92	16.36	<0.001*	0.21	58
ELL	42	45.10	20.45	44.48	50	36.94	18.45	37.21	3.46	0.066	0.37	65
Not FRL	307	63.02	25.59	63.54	298	60.19	22.89	59.65	4.34	0.038*	0.16	56
FRL	465	55.75	23.31	55.26	338	48.45	21.70	49.13	17.19	<0.001*	0.27	61
Male	367	55.28	24.26	55.90	308	52.10	22.97	51.36	7.23	0.007*	0.19	58
Female	405	61.68	24.33	61.40	328	55.69	22.94	56.04	11.21	0.001*	0.23	59

* $p < 0.05$

Houston Independent School District: Results for Spring 2014 PASS Open-Ended and Performance Task

Houston Independent School District (HISD)
Spring 2014 PASS Open-Ended and Performance Task
Key Findings for Phase 1

For all students combined (the "All" group) and the specified subgroups in HISD, the following outcomes favoring Phase 1 Elementary Cohort students were found on the Spring 2014 PASS Performance Task section.

All

- Despite the fact that Phase 2 students had a nearly substantively important (i.e., educationally meaningful) advantage on the pretest (g = -0.24), Phase 1 students scored statistically significantly higher than Phase 2 students, with an effect size (g = 0.39) that was also considered to be substantively important according to WWC standards.

ELL

- Despite the fact that Phase 2 students had an advantage on the pretest (g = -0.14), Phase 1 ELL students scored statistically significantly higher than Phase 2 ELL students with a substantively important effect size (g = 0.38).

Economically Disadvantaged (FRL)

- Phase 1 FRL students scored statistically significantly higher than Phase 2 FRL students with a substantively important effect size (g = 0.40).

Males

- Phase 1 Males scored statistically significantly higher than Phase 2 Males with a substantively important effect size (g = 0.34).

Spring 2014 PASS Performance Task and Open-Ended Results: HISD

A preliminary analysis was conducted on the Spring 2012 OE and PT sections of the PASS to determine baseline equivalence between Phase 1 and Phase 2 students in the elementary cohort included the present analysis (see Table 13) as the PASS OE and PT sections were not administered until the end of the first posttest year, meaning there was no Fall 2011 baseline scores available. 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).

Results indicated that for the elementary cohort aggregate scores (i.e., for all students combined), there was no statistically significant difference or substantively important effect between Phase 1 and Phase 2 students on the Spring 2012 OE percent correct. However, Phase 1 students had a statistically significantly higher mean Spring 2012 PT percent correct compared to Phase 2, with the effect size being substantially important according to WWC standards.

Table 13. PASS OE and PT, Spring 2012 For Students Who Had a Spring 14 OE or PT Score, Treatment (Phase 1) and Control (Phase 2) Means Comparison: HISD

		Treatment (Phase 1)			Control (Phase 2)				
Section	Cohort	n	M	SD	n	M	SD	t	*g*
Open-Ended	Elementary	288	38.19	20.15	174	41.57	17.99	-1.82	-0.17
Performance Task	Elementary	393	54.05	21.09	240	49.24	16.72	3.00*	0.25

* $p < 0.05$

Due to the fact that the PASS OE and PT sections were not administered until the end of the first posttest year, meaning there were no true baseline scores available, and due to a substantively meaningful difference on the Spring 2012 scores, correlation analyses were conducted to examine the relationship between the Spring 2014 PASS OE and PT percent correct and (1) the Spring 2012 PASS OE and PT percent correct, as well as (2) the Fall 2011 PASS Multiple Choice (MC) scaled score results, to determine which scores would serve as the better baseline measure of achievement. The analyses revealed statistically significant, but low correlations among each of the measures of achievement (see Table 14). For the elementary cohort, the correlations for the available measures were both very similar, with the Fall 2011 PASS MC showing a stronger relationship with the Spring 2014 PASS OE and PT percent correct score.

Table 14. Correlations on the Percent Correct for Spring 2014 PASS OE and PT with Spring 2012 PASS OE and PT, and Fall 2011 PASS Multiple Choice: HISD

Spring 2014 PASS	Cohort	Fall 2011 PASS Multiple Choice	Spring 2012 Open-Ended	Spring 2012 Performance Task
Spring 2014 Open-Ended	Elementary	0.38*	0.34*	NA
Spring 2014 Performance Task	Elementary	0.38*	NA	0.36*

* $p < 0.05$

To determine baseline equivalence on the Fall 2011 PASS MC between HISD Phase 1 and Phase 2 students included the present analyses, a series of independent *t*-tests were conducted for all elementary 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 (see Table 15). For the elementary OE cohort in the aggregate (i.e., the "All" group), Phase 1 students again demonstrated a statistically significant advantage over their Phase 2 counterparts in their baseline achievement level ($t(677) = 2.42$, $p = 0.016$, $g = 0.19$, PR = 58), but the effect size linked to this advantage did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$). Consistent with this overall difference in performance, a statistically significant, but not substantively important advantage was observed to favor one subgroup of Phase 1 students in the elementary cohort: Not IEP. In addition, statistically significant and substantively important advantages were observed to favor the Phase 1 IEP, Not ELL, and Female subgroups. Additionally, Not FRL Phase 1 students had an advantage over Phase 2 students that was not statistically significant, but was substantively important ($g = 0.53$).

With respect to the elementary PT cohort in the aggregate (i.e., the "All" group), Phase 1 students demonstrated a statistically significant advantage over their Phase 2 counterparts in their baseline achievement level ($t(698) = 3.12$, $p = 0.002$, $g = 0.24$, PR = 60), but the effect size linked to this advantage did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$). Statistically significant advantages were observed to favor four subgroups of Phase 1 students: namely, Not IEP, Not ELL, Not FRL, and Females, with substantially important effect sizes indicated for the subgroups Not ELL, Not FRL, and Female. Additionally, IEP Phase 1 students had an advantage over Phase 2 students that was not statistically significant, but was substantively important ($g = 0.68$).

While neither the Spring 2012 PASS OE and PT nor the Fall 2011 PASS Multiple Choice provided complete baseline equivalence between Phase 1 and Phase 2 students for all subgroups, the Fall 2011 PASS Multiple Choice was administered as a true baseline assessment vs. the Spring 2012 PASS OE and PT, which was not administered until the end of the first posttest year. Therefore, due to its stronger relationship to the Spring 2014 PASS OE and PT outcomes, and because it was a true baseline measure, the Fall 2011 PASS Multiple Choice scaled score was chosen as the covariate.

Table 15. Fall 2011 PASS Multiple Choice for Students Who Had Spring 2014 OE or PT Scores, Treatment (Phase 1) and Control (Phase 2) Means Comparison: HISD

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort – Open-Ended									
All	404	295.15	100.33	275	276.89	90.27	2.42*	0.19	58
Not IEP	385	297.13	99.99	264	281.22	88.16	2.09*	0.17	57
IEP	19	254.89	101.29	11	173.09	80.95	2.28*	0.84	80
Not ELL	186	326.26	99.38	102	272.84	93.11	4.46*	0.55	71
ELL	218	268.6	93.46	173	279.28	88.75	-1.15	-0.12	45
Not FRL	57	365.16	100.72	14	310.29	106.94	1.81	0.53	70
FRL	347	283.65	95.61	261	275.1	89.18	1.23	0.09	54
Male	212	291.72	104.47	131	284.45	94.39	0.65	0.01	53
Female	192	298.93	95.68	144	270.01	86.11	2.86*	0.31	62
	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort – Performance Task									
All	427	302.74	105.85	273	278.69	88.68	3.12*	0.24	60
Not IEP	408	304.97	105.65	260	283.23	86.5	2.78*	0.22	59
IEP	19	254.89	101.29	13	187.85	86.02	1.95	0.68	75
Not ELL	208	339.5	105.37	108	275.31	89.73	5.39*	0.64	74
ELL	219	267.84	94.02	165	280.9	88.2	-1.38	-0.14	44
Not FRL	76	391.18	105.4	13	315.08	109.73	2.39*	0.71	76
FRL	351	283.6	95.87	260	276.87	87.36	0.89	0.07	53
Male	223	301.26	112.18	129	284.86	92.5	1.41	0.16	56
Female	204	304.36	98.71	144	273.17	85.07	3.07*	0.33	63

* $p < 0.05$

Employing the Fall 2011 PASS MC data as covariates to statistically adjust the outcomes for baseline differences in achievement, preliminary analyses were conducted on Spring 2014 PASS OE and PT percent correct scores to determine any differences between Phase 1 and Phase 2 HISD students in the elementary cohort. As these analyses were exploratory in nature, no corrections were made for multiple comparisons. As noted earlier, the elementary cohort Phase 1 and Phase 2 students were not equivalent on the baseline measure for the OE and PT, with Phase 1 students having an advantage for several subgroups on both the OE and PT. Due to these baseline differences, there were violations of the ANCOVA assumption of equal variances. Therefore, results for these particular groups should be interpreted with the Phase 1 advantage and the statistical issues in mind.

Elementary Cohort PASS Open-Ended Analyses: HISD

With respect to the 679 elementary cohort students in Phase 1 (*n* = 404) and Phase 2 (*n* = 275) schools, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students differed by Phase in their percent correct score on the Spring 2014 OE section of the PASS assessment. In addition to these regressions, a second set of ANCOVA analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (*g*), was also conducted on the outcomes for all students by Phase, as well as for subgroups of these same students, categorized by their Special Education (IEP) status, English Language Learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender.

Elementary Cohort Spring 2014 PASS Open-Ended Results: HISD

Among the 679 elementary cohort students across the district, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS Multiple Choice scaled scores (Block 3) explained 16% of the total variance (R^2) in students' Spring 2014 OE scores (see Table 16). The addition of Phase to the model added only 1% to the variance explained, and Phase did not have a statistically significant impact on the Spring 2014 OE percent correct (β = -0.07, *t* = -1.91, *p* = 0.056).

While the ANCOVA adjusted means presented in Table 17 can be seen to trend higher for Phase 1 students (*n* = 404, Adjusted Mean = 62.98) than for Phase 2 students (*n* = 275, Adjusted Mean = 59.90) overall (i.e., the "All" group) (*F* (1, 672) = 3.66, *p* = 0.056, *g* = 0.14, *PR* = 56), the difference was not statistically significant, nor was the magnitude of the effect size (*g* = 0.14) considered to be substantively important. Two subgroups (Not IEP and FRL) had statistically significant differences favoring Phase 1 students, but the magnitude of the effects were not considered to be substantively important. One additional subgroup (IEP) did not indicate a statistically significant difference between Phase 1 and Phase 2 students, but the effect size (*g* = -0.29) was considered substantively important, with Phase 2 students having the advantage, even though Phase 1 students had a substantively important advantage on the pretest. However, it should be noted that the sample sizes for the IEP group (Phase 1 *n* = 19; Phase 2 *n* = 11) were very small, and therefore may not be representative. In addition, the sample sizes for the Not FRL group (Phase 1 *n* = 54; Phase 2 *n* = 14) were unbalanced and very small for the Phase 2 group, and also may not be representative.

Table 16. PASS Open-Ended Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Elementary Cohort (N = 679): HISD

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 674) = 4.40$, $p = .002$, $R^2 = .03$, F Change $(4, 674) = 4.40$, $p = .002$)					
IEP (0 = No, 1 = IEP)	-11.90	4.09	-0.11	-2.91	0.004*
ELL (0 = No, 1 = ELL)	-2.81	1.77	-0.06	-1.58	0.114
FRL (0 = No, 1 = FRL)	-3.43	2.86	-0.05	-1.20	0.231
Gender (0 = M, 1= F)	3.37	1.68	0.08	2.01	0.045*
Block 2: Demographics + Fall 2011 PASS Score Model Fit: $F(5, 673) = 24.51$, $p < .001$, $R^2 = .15$, F Change $(1, 673) = 102.32$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-6.04	3.86	-0.06	-1.57	0.117
ELL (0 = No, 1 = ELL)	-0.71	1.67	-0.02	-0.43	0.668
FRL (0 = No, 1 = FRL)	1.97	2.72	0.03	0.72	0.469
Gender (0 = M, 1= F)	3.49	1.57	0.08	2.22	0.026*
Fall 2011 PASS Scaled Score	0.09	0.01	0.38	10.12	<0.001*
Block 3: Demographics + Fall 2011 PASS Score + Phase Model Fit: $F(6, 672) = 21.12$, $p < .001$, $R^2 = .16$, F Change $(1, 672) = 3.66$, $p = .056$)					
IEP (0 = No, 1 = IEP)	-6.21	3.85	-0.06	-1.61	0.107
ELL (0 = No, 1 = ELL)	-0.57	1.66	-0.01	-0.35	0.730
FRL (0 = No, 1 = FRL)	2.54	2.73	0.04	0.93	0.352
Gender (0 = M, 1= F)	3.61	1.57	0.08	2.31	0.021*
Fall 2011 PASS Scaled Score	0.09	0.01	0.37	10.01	<0.001*
Phase (1= P1, 2 = P2)	-3.09	1.61	-0.07	-1.91	0.056

$p < 0.05$

Table 17. PASS Open-Ended Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Elementary Cohort (N = 679): HISD

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	404	63.45	22.32	62.98	275	59.21	21.50	59.90	3.66	0.056	0.14	56
Not IEP	385	64.03	21.84	63.64	264	59.66	21.47	60.23	4.34	0.038*	0.16	56
IEP	19	51.75	28.81	47.68	11	48.48	20.35	55.53	0.70	0.410	-0.29	38
Not ELL	186	65.68	21.66	64.32	102	59.80	23.38	62.30	0.56	0.455	0.09	54
ELL	218	61.54	22.75	62.10	173	58.86	20.38	58.16	3.71	0.055	0.18	57
Not FRL	57	66.67	20.41	66.04	14	61.90	23.96	64.47	0.06	0.803	0.07	53
FRL	347	62.92	22.60	62.71	261	59.07	21.40	59.35	4.02	0.045*	0.15	56
Male	212	61.71	23.47	61.44	131	57.38	22.57	57.82	2.26	0.134	0.16	56
Female	192	65.36	20.88	64.67	144	60.88	20.42	61.81	1.72	0.191	0.14	55

* $p < 0.05$

Elementary Cohort PASS Performance Task Analyses: HISD

With respect to the 700 elementary cohort students in Phase 1 (n = 427) and Phase 2 (n = 273) schools, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students differed by Phase in their percent correct score on the Spring 2014 PT section of the PASS assessment. In addition to these regressions, a second set of ANCOVA analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (g), was also conducted on the outcomes for all students by Phase, as well as for subgroups of these same students, categorized by their Special Education (IEP) status, English Language Learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender.

Elementary Cohort Spring 2014 PASS Performance Task Results: HISD

Among the 700 elementary cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS Multiple Choice scaled scores (Block 3) explained 22% of the total variance (R^2) in students' Spring 2014 PT scores (see Table 18). While the addition of Phase to the model added only 4% to the variance explained, and there was a statistically significant difference in the Spring 2014 PT percent correct favoring Phase 1 students (β = -0.18, t = -5.37, $p < 0.001$).

The ANCOVA adjusted means presented in Table 19 can be seen to be statistically significantly higher for Phase 1 students (n = 427, Adjusted Mean = 65.47) than for Phase 2 students (n = 273, Adjusted Mean = 59.18) overall (i.e., the "All" group) (F (1, 693) = 28.84, $p < 0.001$, g = 0.39, PR = 65), indicating that the average Phase 1 student scored at the 65th percentile of the control group. The effect size (g = 0.39) was also considered to be substantively important according to WWC standards. Consistent with these overall outcomes, four subgroups (Not IEP, ELL, FRL, and Males) were statistically significant and favored Phase 1 students, along with substantively important effect sizes (see Table 19).For the ELL subgroup, this positive finding on the posttest appeared despite the fact that Phase 2 students had an advantage on the pretest (g = -0.14).

Table 18. PASS Performance Task Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Elementary Cohort (N = 700): HISD

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 695) = 16.68$, $p < .001$, $R^2 = .09$, *F* Change (4, 695) = 16.68, $p < .001$)					
IEP (0 = No, 1 = IEP)	-17.66	2.90	-0.22	-6.10	<0.001*
ELL (0 = No, 1 = ELL)	-0.39	1.29	-0.01	-0.30	0.761
FRL (0 = No, 1 = FRL)	-8.94	1.92	-0.18	-4.65	<0.001*
Gender (0 = M, 1= F)	2.10	1.21	0.06	1.73	0.083
Block 2: Demographics + Fall PASS MC Score Model Fit: $F(5, 694) = 31.34$, $p < .001$, $R^2 = .18$, *F* Change (1, 694) = 82.20, $p < .001$)					
IEP (0 = No, 1 = IEP)	-13.85	2.77	-0.17	-5.00	<0.001*
ELL (0 = No, 1 = ELL)	1.09	1.23	0.03	0.89	0.373
FRL (0 = No, 1 = FRL)	-4.17	1.89	-0.08	-2.20	0.028*
Gender (0 = M, 1= F)	2.22	1.15	0.07	1.93	0.054
Fall PASS MC Scaled Score	0.06	0.01	0.34	9.07	<0.001*
Block 3: Demographics + Fall PASS MC Score + Phase Model Fit: $F(6, 693) = 31.97$, $p < .001$, $R^2 = .22$, *F* Change (1, 693) = 28.84, $p < .001$)					
IEP (0 = No, 1 = IEP)	-13.88	2.72	-0.17	-5.11	<0.001*
ELL (0 = No, 1 = ELL)	1.25	1.20	0.04	1.04	0.300
FRL (0 = No, 1 = FRL)	-2.68	1.88	-0.05	-1.43	0.154
Gender (0 = M, 1= F)	2.46	1.12	0.07	2.19	0.029*
Fall PASS MC Scaled Score	0.05	0.01	0.33	8.94	<0.001*
Phase (1= P1, 2 = P2)	-6.30	1.17	-0.18	-5.37	<0.001*

* $p < 0.05$

Table 19. PASS Performance Task Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Elementary Cohort (N = 700): HISD

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	427	66.04	14.67	65.47	273	58.28	18.45	59.18	28.84	<0.001*	0.39	65
Not IEP	408	66.62	14.31	66.13	260	59.48	17.35	60.25	24.63	<0.001*	0.38	65
IEP	19	53.56	17.08	50.35	13	34.39	23.75	39.08	2.43	0.131	0.55	71
Not ELL	208	67.87	15.22	66.52	108	57.08	19.19	59.68	12.88	<0.001*	0.41	66
ELL	219	64.30	13.94	64.62	165	59.07	17.96	58.66	15.48	<0.001*	0.38	65
Not FRL	76	72.21	13.40	71.70	13	64.25	17.23	67.25	1.19	0.278	0.32	62
FRL	351	64.71	14.61	64.61	260	57.99	18.49	58.12	27.42	<0.001*	0.40	65
Male	223	64.73	16.19	64.18	129	57.36	18.43	58.32	11.23	0.001*	0.34	63
Female	204	67.47	12.69	66.77	144	59.11	18.49	60.11	17.46	<0.001*	0.43	67

* $p < 0.05$

New Mexico Region: Results for Spring 2014 PASS Open-Ended and Performance Task

New Mexico
Fall 2014 Annual Report Open-Ended and Performance Task
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 PASS Open-Ended and Performance Task sections.

ELL

- Elementary Cohort Open-Ended: Phase 1 ELL students had a substantively important effect size (g = 0.36).
- Elementary Cohort Performance Task: Phase 1 ELL students scored statistically significantly higher than Phase 2, with an effect size (g = 0.52) that was substantively important.

Females

- Elementary Cohort Performance Task: While Phase 2 females had a nearly substantively important advantage on the pretest (g = -0.24), Phase 1 females scored higher on the posttest, although the difference was not substantively important (g =0.14).

IEP

- Elementary Cohort Open-Ended: After controlling for the substantively important advantage Phase 2 IEP students had on the pretest (g = -0.26), Phase 1 IEP students scored higher on the posttest, although the difference was not substantively important (g = 0.16). The sample size for Phase 2 (n = 20), however, was small.
- Elementary Cohort Performance Task: Phase 1 IEP students scored statistically significantly higher than Phase 2, with an effect size (g = 0.47) that was substantively important. However, the sample size for Phase 2 (n = 22) was small.

Males

- Elementary Cohort Performance Task: Phase 1 Males scored statistically significantly higher than Phase 2, with an effect size (g = 0.24) that nearly reached the substantively important threshold.

Spring 2014 PASS Performance Task and Open-Ended Results: New Mexico

A preliminary analysis was conducted on the Spring 2012 OE and PT sections of the PASS to determine baseline equivalence between Phase 1 and Phase 2 students in the elementary and middle school cohorts included the present analysis (see Table 20) as the PASS OE and PT sections were not administered until the end of the first posttest year, meaning there was no Fall 2011 baseline scores available. 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).

Results indicated that for the elementary cohort aggregate scores (i.e., for all students combined), there were no statistically significant differences in the Spring 2012 OE or PT percent correct scores between Phase 1 and Phase 2 students, and no substantially important effect sizes according to WWC standards. However, for the middle school cohort aggregate scores, Phase 1 students had statistically significantly higher mean Spring 2012 OE and Spring 2012 PT percent correct scores, with the magnitude of both effects being substantially important.

Table 20. PASS OE and PT, Spring 2012 For Students Who Had a Spring 14 OE or PT Score, Treatment (Phase 1) and Control (Phase 2) Means Comparison: New Mexico

		Treatment (Phase 1)			Control (Phase 2)				
Section	Cohort	n	M	SD	n	M	SD	t	*g*
Open-Ended	Elementary	328	49.24	21.14	175	47.33	20.2	0.98	0.09
Performance Task	Elementary	329	58.31	19.66	182	56.53	16.55	1.03	0.1
Open-Ended	Middle School	426	72.24	16.63	112	64.76	16.39	4.25*	0.45
Performance Task	Middle School	400	55.01	19.51	75	41.25	25.12	5.34*	0.67

* $p < 0.05$

Due to the fact that the PASS OE and PT were not administered until the end of the first posttest year, meaning there were no true baseline scores available, and due to substantively meaningful differences on the Spring 2012 scores, correlation analyses were conducted to examine the relationship between the Spring 2014 PASS OE and PT percent correct and (1) the Spring 2012 PASS OE and PT percent correct, as well as (2) the Fall 2011 PASS Multiple Choice (MC) scaled score results, to determine which scores would serve as the better baseline measure of achievement. The analyses revealed statistically significant, but low correlations among each of the measures of achievement (see Table 21). For both the elementary and middle school cohorts, the Fall 2011 PASS MC scaled scores had higher statistically significant correlations with the Spring 2014 PASS OE and PT.

Table 21. Correlations on the Percent Correct for Spring 2014 PASS OE and PT with Spring 2012 PASS OE and PT, and Fall 2011 PASS Multiple Choice: New Mexico

Spring 2014 PASS	Cohort	Fall 2011 PASS Multiple Choice	Spring 2012 Open-Ended	Spring 2012 Performance Task
Spring 2014 Open-Ended	Elementary	0.36*	0.31*	NA
	Middle School	0.45*	0.38*	NA
Spring 2014 Performance Task	Elementary	0.37*	NA	0.33*
	Middle School	0.43*	NA	0.42*

* $p < 0.05$

To determine baseline equivalence on the Fall 2011 PASS MC between New Mexico Phase 1 and Phase 2 students included the present analyses, a series of independent *t*-tests were 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 (see Table 22). For the elementary OE cohort in the aggregate (i.e., the "All" group), there was no statistically significant difference between Phase 1 and Phase 2 in their baseline achievement levels ($t(574) = -1.55$, $p = 0.123$, $g = -0.14$, PR = 45), and the effect size did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$). Consistent with this overall difference in performance, no statistically significant differences between Phase 1 and Phase 2 students were observed for any of the eight subgroups. Among elementary cohort students who were classified as IEP, however, while the advantage seen for those in Phase 2 proved not to be statistically significant, it was substantively important. The IEP sample size for Phase 2 ($n = 20$), was small and less than half the size of Phase 1 ($n = 50$), and therefore may not be representative. For students in the middle school OE cohort, a statistically significant difference in aggregate performance (i.e., the "All" group) for Phase 1 students was observed ($t(554) = 11.01$, $p < 0.001$, $g = 1.16$, PR = 88), with an effect size linked to this difference that exceed the WWC criteria for substantive importance. However, the sample size for Phase 1 ($n = 443$) was nearly four times as large as that for Phase 2 ($n = 113$). As a result, the baseline equivalence results may not be representative of Phase 2 performance. Statistically significant and substantively important advantages in baseline performance for Phase 1 middle school students were also observed for all eight subgroups. Like with the overall (All) group, however, the sample sizes for Phase 1 were (with the exception of the ELL subgroup, which was small ($n = 16$) and half the size of Phase 2 ($n = 33$)) much larger than Phase 2, ranging from twice (IEP) to over 30 times as large (Not FRL) as Phase 2. As a result, the baseline outcomes may not be representative.

With respect to the elementary PT cohort in the aggregate (i.e., the "All" group), there was no statistically significant difference between Phase 1 and Phase 2 in their baseline achievement levels ($t(571) = -1.69$, $p = 0.093$, $g = -0.15$, PR = 44), and the effect size did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$). Consistent with this overall difference in performance, no statistically significant or substantively important advantages were observed for any of the eight subgroups. With respect to students in the middle school PT cohort, a statistically significant difference in aggregate performance (i.e., the "All" group) for Phase 1 students was observed ($t(519) = 9.05$, $p < 0.001$, $g = 1.06$, PR = 85), linked to an effect size that exceeded the WWC criteria for substantive importance. However, the sample size for Phase 1 ($n = 443$) was nearly five times as large as that for Phase 2 ($n = 88$). As a result, the baseline equivalence results may not be representative of Phase 2 performance. Statistically significant and substantively important advantages in baseline performance were also observed for all but one subgroup of Phase 1 middle school cohort students (Not FRL). Among middle school cohort students who were classified as Not FRL, while the advantage seen for those in Phase 1 proved not to be statistically significant, it was substantively important. Like with the overall (All) group, however, the sample sizes for Phase 1 were (with the exception of the ELL subgroup, which was small ($n = 16$) and

three-quarters the size of Phase 2 (*n* = 22)) much larger than Phase 2, ranging from nearly three times (IEP) to nearly 30 times as large (Not FRL) as Phase 2. As a result, the baseline outcomes may not be representative.

While neither the Spring 2012 PASS OE and PT nor the Fall 2011 PASS Multiple Choice provided complete baseline equivalence between Phase 1 and Phase 2 students either for all students combined or for all subgroups, the Fall 2011 PASS Multiple Choice was administered as a true baseline assessment vs. the Spring 2012 PASS OE and PT, which was not administered until the end of the first posttest year. Therefore, due to its stronger relationship to the Spring 2014 PASS OE and PT outcomes, and because it was a true baseline measure, the Fall 2011 PASS Multiple Choice scaled score was chosen as the covariate for both the elementary and middle school cohorts.

Table 22. Fall 2011 PASS Multiple Choice for Students Who Had Spring 2014 PT or OE Scores, Treatment (Phase 1) and Control (Phase 2) Means Comparison: New Mexico

Group	Treatment (Phase 1)			Control (Phase 2)			*t*	*g*	PR
	n	*M*	*SD*	*n*	*M*	*SD*			
Elementary Cohort – Open-Ended									
All	383	320.17	106.42	193	335.15	116.01	-1.55	-0.14	45
Not IEP	333	325.74	106.65	173	338.02	115.34	-1.19	-0.11	46
IEP	50	283.08	98.04	20	310.35	121.83	-0.98	-0.26	40
Not ELL	319	332.72	103.62	161	348.68	115.17	-1.53	-0.15	44
ELL	64	257.64	98.5	32	267.09	95.73	-0.45	-0.1	46
Not FRL	190	358.88	97.33	106	365	117.01	-0.48	-0.06	48
FRL	193	282.07	101.32	87	298.78	104.37	-1.27	-0.16	44
Male	193	328.99	111.97	95	337.81	115.22	-0.62	-0.08	47
Female	190	311.22	99.96	98	332.57	117.3	-1.62	-0.2	42
Group	Treatment (Phase 1)			Control (Phase 2)			*t*	*g*	PR
	n	*M*	*SD*	*n*	*M*	*SD*			
Middle School Cohort – Open-Ended									
All	443	377.52	99.54	113	260.04	107.59	11.01*	1.16	88
Not IEP	401	387.97	95.41	91	276.53	101.51	9.94*	1.15	88
IEP	42	277.83	82.1	22	191.86	107.39	3.57*	0.93	82
Not ELL	427	380.96	98.28	80	280.41	111.64	8.21*	1	84
ELL	16	285.94	91.34	33	210.67	78.71	2.98*	0.89	81
Not FRL	182	409.77	92.88	6	331.17	149.72	2.00*	0.83	80
FRL	261	355.03	97.99	107	256.06	104.27	8.64*	0.99	84
Male	205	363.21	104.73	56	254.63	118.06	6.69*	1.01	84
Female	238	389.85	93.31	57	265.37	96.98	8.98*	1.32	91
Group	Treatment (Phase 1)			Control (Phase 2)			*t*	*g*	PR
	n	*M*	*SD*	*n*	*M*	*SD*			
Elementary Cohort – Performance Task									
All	376	320.14	105.68	197	336.32	115.65	-1.69	-0.15	44
Not IEP	328	325.03	106.31	175	342.31	114.07	-1.69	-0.16	44
IEP	48	286.69	95.71	22	288.68	119.83	-0.08	-0.02	49
Not ELL	312	332.63	102.75	166	348.51	115.37	-1.54	-0.15	44
ELL	64	259.25	99.06	31	271.03	94.64	-0.55	-0.12	45
Not FRL	183	358.82	97.49	112	362.98	117.97	-0.33	-0.04	48
FRL	193	283.46	100.12	85	301.19	103.06	-1.35	-0.18	43
Male	190	328.19	111.48	99	336.06	118.29	-0.56	-0.07	47
Female	186	311.9	99.03	98	336.58	113.54	-1.9	-0.24	41

Table 22, continued

Group	Treatment (Phase 1)			Control (Phase 2)			*t*	*g*	PR
	n	*M*	*SD*	*n*	*M*	*SD*			
Middle School Cohort – Performance Task									
All	433	377.88	98.76	88	271.13	110.76	9.05*	1.06	85
Not IEP	392	388.11	94.78	73	285.16	106.33	8.35*	1.06	86
IEP	41	280.07	81.81	15	202.8	109.81	2.85*	0.85	80
Not ELL	417	381.41	97.42	66	290.95	113.07	6.85*	0.91	82
ELL	16	285.94	91.34	22	211.64	79.42	2.67*	0.86	81
Not FRL	178	408.52	92.77	6	331.17	149.72	1.97	0.81	79
FRL	255	356.49	97.34	82	266.73	107.27	7.08*	0.9	82
Male	201	364.68	104.14	41	272.07	124.05	5.02*	0.86	80
Female	232	389.32	92.56	47	270.3	99.09	7.94*	1.27	90

* $p < 0.05$

Employing the Fall 2011 PASS MC data as covariates to statistically adjust the outcomes for baseline differences in achievement, preliminary analyses were conducted on Spring 2014 PASS OE and PT percent correct scores to determine any differences between Phase 1 and Phase 2 New Mexico students in the elementary and middle school cohorts. As these analyses were exploratory in nature, no corrections were made for multiple comparisons. As noted earlier, for both the elementary and middle school cohorts, Phase 1 and Phase 2 students were not equivalent on the baseline measure for the OE and PT, with Phase 1 students having an advantage both overall (for the middle school OE and PT), and for several subgroups on both the OE and PT. Due to these baseline differences, there were violations of the ANCOVA assumption of equal variances. Therefore, results for these particular groups should be interpreted with the Phase 1 advantage, the statistical issues, and some of the large differences in samples sizes in mind.

Elementary and Middle School Cohorts PASS Open-Ended Analyses: New Mexico

With respect to the 576 elementary cohort students in Phase 1 (n = 383) and Phase 2 (n = 193) schools and the 556 middle school cohort students in Phase 1 (n = 443) and Phase 2 (n =113) schools, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their percent correct score on the Spring 2014 OE section of the PASS assessment. In addition to these regressions, a second set of ANCOVA analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (g), was also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their Special Education (IEP) status, English Language Learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender.

Elementary Cohort Spring 2014 PASS Open-Ended Results: New Mexico

Among the 576 elementary cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS Multiple Choice scaled scores (Block 3) explained 17% of the total variance (R^2) in students' Spring 2014 OE scores (see Table 23. Adding Phase to the model did not increase the variance explained, and Phase did not have a statistically significant impact on the Spring 2014 OE percent correct (β = -0.05, t = -1.34, p = 0.182).

While the ANCOVA adjusted means presented in Table 24 can be seen to trend higher for Phase 1 students (n = 383, Adjusted Mean = 68.54) than for Phase 2 students (n = 193, Adjusted Mean = 66.41) overall (i.e., the "All" group) (F (1, 569) = 1.78, p = 0.182, g = 0.11, PR = 54), the difference was not

statistically significant, nor was the magnitude of the effect size (g = 0.11) considered to be substantively important. No subgroup comparison indicated a statistically significant difference between the adjusted means for Phase 1 and Phase 2. Only one subgroup, ELL (g = 0.36) had an effect size that was considered to be substantively important according to WWC guidelines, and favored Phase 1. In addition, after controlling for the substantively important advantage Phase 2 IEP students had on the pretest (g = -0.26), there was no substantively important difference between the groups on the posttest (g = 0.16). Moreover, Phase 1 students demonstrated a non-statistically significant positive achievement advantage on the posttest in each subgroup (g = 0.06 - g = 0.36) over Phase 2 students after controlling for the Phase 2 baseline non-statistically significant achievement advantage for every subgroup.

Table 23. PASS Open-Ended Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Elementary Cohort (N = 576): New Mexico

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 571) = 13.80$, $p < .001$, $R^2 = .09$, *F* Change (4, 571) = 13.80, $p < .001$)					
IEP (0 = No, 1 = IEP)	-11.31	2.44	-0.19	-4.63	<0.001*
ELL (0 = No, 1 = ELL)	-8.76	2.18	-0.17	-4.01	<0.001*
FRL (0 = No, 1 = FRL)	-4.95	1.65	-0.13	-3.01	0.003*
Gender (0 = M, 1= F)	2.82	1.58	0.07	1.79	0.074
Block 2: Demographics + Fall 2011 PASS Score Model Fit: $F(5, 570) = 22.66$, $p < .001$, $R^2 = .17$, *F* Change (1, 570) = 53.04, $p < .001$)					
IEP (0 = No, 1 = IEP)	-8.04	2.38	-0.13	-3.38	0.001*
ELL (0 = No, 1 = ELL)	-5.73	2.13	-0.11	-2.69	0.007*
FRL (0 = No, 1 = FRL)	-1.19	1.66	-0.03	-0.72	0.472
Gender (0 = M, 1= F)	3.69	1.52	0.09	2.43	0.015*
Fall 2011 PASS Scaled Score	0.06	0.01	0.31	7.28	<0.001*
Block 3: Demographics + Fall 2011 PASS Score + Phase Model Fit: $F(6, 569) = 19.20$, $p < .001$, $R^2 = .17$, *F* Change (1, 569) = 1.78, $p = .182$)					
IEP (0 = No, 1 = IEP)	-8.15	2.38	-0.14	-3.42	0.001*
ELL (0 = No, 1 = ELL)	-5.67	2.13	-0.11	-2.67	0.008*
FRL (0 = No, 1 = FRL)	-1.28	1.66	-0.03	-0.77	0.439
Gender (0 = M, 1= F)	3.71	1.51	0.94	2.45	0.015*
Fall 2011 PASS Scaled Score	0.06	0.01	0.31	7.34	<0.001*
Phase (1= P1, 2 = P2)	-2.13	1.60	-0.05	-1.34	0.182

$p < 0.05$

Table 24. PASS Open-Ended Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Elementary Cohort (N = 576): New Mexico

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	383	68.15	19.31	68.54	193	67.18	20.33	66.41	1.78	0.182	0.11	54
Not IEP	333	69.52	19.00	69.76	173	68.21	19.87	67.74	1.42	0.233	0.10	54
IEP	50	59.00	19.11	59.71	20	58.33	22.62	56.55	0.46	0.502	0.16	56
Not ELL	319	69.44	18.84	69.87	161	69.67	18.90	68.82	0.38	0.538	0.06	52
ELL	64	61.72	20.51	61.95	32	54.69	22.89	54.22	2.98	0.088	0.36	64
Not FRL	190	70.61	17.58	70.85	106	70.28	19.25	69.86	0.23	0.634	0.02	51
FRL	193	65.72	20.64	66.13	87	63.41	21.09	62.49	2.25	0.135	0.17	57
Male	193	66.32	19.32	66.49	95	65.61	20.02	65.26	0.30	0.587	0.06	53
Female	190	70.00	19.18	70.56	98	68.71	20.62	67.62	1.69	0.195	0.15	56

* $p < 0.05$.

Middle School Cohort Spring 2014 PASS Open-Ended Results: New Mexico

Among the 556 middle school cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS MC scaled scores (Block 3) explained 29% of the total variance (R^2) in students' 2014 Spring OE scores. Adding Phase to the model did not increase the variance explained, and Phase did not have a statistically significant impact on the Spring 2014 OE percent correct, (β = -0.03, t = -0.71, p = 0.477) (see Table 25).

While Phase 1 students had substantively important advantages on the pretest for all groups, none of the differences for any groups on the posttest were statistically significant. Therefore, while the ANCOVA adjusted means presented in Table 26 indicate no statistically significant difference between Phase 1 students (n = 443, Adjusted Mean = 85.98) and Phase 2 students (n = 113, Adjusted Mean = 84.82) overall (i.e., the "All" group) (F (1, 549) = 0.51, p = 0.477, g = 0.08, PR = 53), the Phase 1 group had a substantively meaningful advantage on the pretest. While not statistically significant, the effect size for the IEP subgroup was substantively important, favoring Phase 2 students. It should be noted, however, that the same size for Phase 2 IEP students (n = 22) was small, and therefore may not be representative.

It should also be noted that the overall (All) sample size for Phase 1 (n = 443) was nearly four times as large as that for Phase 2 (n = 113). Furthermore, like with the All group, the remaining sample sizes for Phase 1 were (with the exception of the ELL subgroup, which was small (n = 16) and half the size of Phase 2 (n = 33)) much larger than Phase 2, ranging from twice (IEP) to over 30 times as large (Not FRL) as Phase 2. As a result, the posttest outcomes may not be representative.

Table 25. PASS Open-Ended Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Middle School Cohort (N = 556): New Mexico

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 551) = 47.83$, $p < .001$, $R^2 = .22$, F Change $(4, 551) = 37.83$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-13.11	1.96	-0.27	-6.69	<0.001*
ELL (0 = No, 1 = ELL)	-12.33	2.21	-0.22	-5.57	<0.001*
FRL (0 = No, 1 = FRL)	-3.83	1.28	-0.12	-2.98	0.003*
Gender (0 = M, 1= F)	3.58	1.22	0.11	2.93	0.003*
Block 2: Demographics + Fall 2011 PASS MC Score Model Fit: $F(5, 550) = 44.56$, $p < .001$, $R^2 = .29$, F Change $(1, 550) = 56.29$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-9.25	1.94	-0.19	-4.77	<0.001*
ELL (0 = No, 1 = ELL)	-8.97	2.16	-0.16	-4.16	<0.001*
FRL (0 = No, 1 = FRL)	-0.95	1.28	-0.03	-0.74	0.460
Gender (0 = M, 1= F)	3.10	1.16	0.10	2.66	0.008*
Fall 2011 PASS MC SS	0.05	0.01	0.31	7.50	<0.001*
Block 3: Demographics + Fall 2011 PASS MC Score + Phase Model Fit: $F(6, 549) = 37.18$, $p < .001$, $R^2 = .29$, F Change $(1, 549) = 0.51$, $p = .477$)					
IEP (0 = No, 1 = IEP)	-9.33	1.94	-0.19	-4.80	<0.001*
ELL (0 = No, 1 = ELL)	-8.57	2.23	-0.16	-3.84	<0.001*
FRL (0 = No, 1 = FRL)	-0.81	1.30	-0.02	-0.62	0.536
Gender (0 = M, 1= F)	3.12	1.17	0.10	2.68	0.008*
Fall 2011 PASS MC SS	0.04	0.01	0.30	6.94	<0.001*
Phase (1= P1, 2 = P2)	-1.15	1.62	-0.03	-0.71	0.477

$p < 0.05$

Table 26. PASS Open-Ended Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Middle School Cohort (N = 556): New Mexico

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	443	87.71	14.57	85.98	113	78.05	17.31	84.82	0.51	0.477	0.08	53
Not IEP	401	89.36	12.65	88.24	91	80.88	12.97	85.80	2.41	0.121	0.19	58
IEP	42	71.90	21.11	67.46	22	66.36	26.47	74.85	1.13	0.291	-0.32	38
Not ELL	427	88.40	13.92	87.62	80	81.83	12.32	86.02	0.97	0.325	0.12	55
ELL	16	69.17	19.46	69.85	33	68.89	23.49	68.56	0.03	0.861	0.06	52
Not FRL	182	89.85	12.01	89.73	6	87.78	11.48	91.41	0.13	0.718	-0.14	44
FRL	261	86.21	15.97	84.01	107	77.51	17.46	82.87	0.37	0.541	0.07	53
Male	205	84.65	16.81	82.49	56	74.64	21.44	82.55	0.00	0.983	0.00	50
Female	238	90.34	11.74	89.10	57	81.40	11.18	86.55	1.82	0.179	0.22	59

* $p < 0.05$

Elementary and Middle School Cohorts PASS Performance Task Analyses: New Mexico

With respect to the 573 elementary cohort students in Phase 1 (*n* = 376) and Phase 2 (*n* = 197) schools and the 521 middle school cohort students in Phase 1 (*n* = 433) and Phase 2 (*n* =88) schools, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their percent correct score on the Spring 2014 PT section of the PASS assessment. In addition to these regressions, a second set of ANCOVA analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (*g*), was also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their Special Education (IEP) status, English Language Learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender.

Elementary Cohort Spring 2014 PASS Performance Task Results: New Mexico

Among the 573 elementary cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS Multiple Choice scaled scores (Block 3) explained 17% of the total variance (R^2) in students' Spring 2014 PT scores (see Table 27). While the addition of Phase to the model did not increase the variance explained, there was a statistically significant difference in the Spring 2014 PT percent correct favoring Phase 1 students (β = -0.09, t = -2.40, p = = 0.017).

The ANCOVA adjusted means presented in Table 28 can be seen to be statistically significantly higher for Phase 1 students (*n* = 376, Adjusted Mean = 68.33) than for Phase 2 students (*n* = 197, Adjusted Mean = 65.78) overall (i.e., the "All" group) (*F* (1, 655) = 5.78, *p* = 0.017, *g* = 0.19, *PR* = 58), indicating that the average Phase 1 student scored at the 58th percentile of the control group. However, the effect size was not considered to be substantively important according to WWC standards. Three subgroup analyses (ELL, IEP, and Males) were statistically significant with differences favoring Phase 1 elementary cohort students, as indicated by the positively signed effect sizes (see Table 28). Only two of these subgroups (ELL *g* = 0.52 and IEP *g* = 0.47) had an effect size that was considered to be substantively important, with the effect size for Males (*g* = 0.24) nearly reaching the substantively important threshold. In addition, while Phase 2 Females had a nearly substantively important advantage on the pretest (g = -0.24), there was no substantively important difference between the groups on the posttest (*g* =0.14). It should also be noted that even though Phase 2 students had higher, non-statistically significant scores on the pretest overall and for all subgroups (as indicated by the negatively signed effect sizes in Table 22), after controlling for pretest differences, Phase 1 students outperformed Phase 2 students on the posttest for all groups.

Table 27. PASS Performance Task Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Elementary Cohort (N = 573): New Mexico

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 568) = 12.42$, $p < .001$, $R^2 = .08$, F Change $(4, 568) = 12.42$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-5.79	2.11	-0.11	-2.75	0.006*
ELL (0 = No, 1 = ELL)	-7.36	1.89	-0.16	-3.90	<0.001*
FRL (0 = No, 1 = FRL)	-5.32	1.42	-0.16	-3.75	<0.001*
Gender (0 = M, 1= F)	3.08	1.37	0.09	2.25	0.025*
Block 2: Demographics + Fall PASS MC Score Model Fit: $F(5, 567) = 22.43$, $p < .001$, $R^2 = .17$, F Change $(1, 567) = 57.52$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-2.65	2.05	-0.05	-1.29	0.197
ELL (0 = No, 1 = ELL)	-4.72	1.83	-0.10	-2.57	0.010*
FRL (0 = No, 1 = FRL)	-2.02	1.42	-0.06	-1.42	0.155
Gender (0 = M, 1= F)	3.71	1.31	0.11	2.84	0.005*
Fall PASS MC Scaled Score	0.05	0.01	0.32	7.58	<0.001*
Block 3: Demographics + Fall PASS MC Score + Phase Model Fit: $F(6, 566) = 19.81$, $p < .001$, $R^2 = .17$, F Change $(1, 566) = 5.78$, $p = .017$)					
IEP (0 = No, 1 = IEP)	-2.77	2.04	-0.05	-1.36	0.175
ELL (0 = No, 1 = ELL)	-4.66	1.83	-0.10	-2.55	0.011*
FRL (0 = No, 1 = FRL)	-2.24	1.42	-0.07	-1.58	0.115*
Gender (0 = M, 1= F)	3.73	1.30	0.11	2.87	0.004*
Fall PASS MC Scaled Score	0.05	0.01	0.33	7.71	<0.001*
Phase (1= P1, 2 = P2)	-3.27	1.36	-0.09	-2.40	0.017*

* $p < 0.05$

Table 28. PASS Performance Task Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Elementary Cohort (N = 573): New Mexico

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	376	67.94	16.50	68.33	197	65.78	17.45	65.05	5.78	0.017*	0.19	58
Not IEP	328	68.35	16.83	68.72	175	66.76	16.67	66.06	3.32	0.069	0.16	56
IEP	48	65.20	13.93	65.43	22	58.02	21.65	57.51	4.81	0.032*	0.47	68
Not ELL	312	69.10	16.46	69.43	166	67.93	16.15	67.30	2.07	0.151	0.13	55
ELL	64	62.32	15.64	62.60	31	54.27	19.84	53.69	7.51	0.007*	0.52	70
Not FRL	183	71.26	15.54	71.37	112	68.33	16.15	68.15	3.26	0.072	0.20	58
FRL	193	64.80	16.80	65.11	85	62.42	18.60	61.70	2.76	0.098	0.20	58
Male	190	66.87	17.33	67.01	99	63.04	18.96	62.78	4.23	0.041*	0.24	59
Female	186	69.04	15.58	69.61	98	68.55	15.39	67.47	1.48	0.225	0.14	55

* $p < 0.05$

Middle School Cohort Spring 2014 PASS Performance Task Results: New Mexico

Among the 521 middle school cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS MC scaled scores (Block 3) explained 27% of the total variance (R^2) in students' 2014 Spring PT scores (see Table 29). While the addition of Phase to the model only increased the variance explained by 3%, Phase had a statistically significant impact on the Spring 2014 PT percent correct, favoring Phase 1 students (β = -0.19, t = -4.45, $p < 0.001$).

While the he ANCOVA adjusted means presented in Table 30 can be seen to be statistically significantly higher for Phase 1 students (n = 433, Adjusted Mean = 65.33) than for Phase 2 students (n = 88, Adjusted Mean = 53.97) overall (i.e., the "All" group) (F (1, 514) = 19.81, $p < 0.001$, g = 0.84, PR = 80), indicating that the average Phase 1 student scored at the 80th percentile of the control group, the Phase 1 group had a substantively meaningful advantage on the pretest for all groups.

It should also be noted that the overall (All) sample size for Phase 1 (n = 443) was nearly five times as large as that for Phase 2 (n = 88). Furthermore, like with the All group, the remaining sample sizes for Phase 1 were (with the exception of the ELL subgroup, which was small (n = 16) and three-quarters the size of Phase 2 (n = 22)) much larger than Phase 2, ranging from nearly three times (IEP) to nearly 30 times as large (Not FRL) as Phase 2. As a result, the posttest outcomes may not be representative.

Table 29. PASS Performance Task Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Middle School Cohort (N = 521): New Mexico

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 516) = 23.62$, $p < .001$, $R^2 = .16$, F Change $(4, 516) = 23.62$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-18.54	3.12	-0.25	-5.94	<0.001*
ELL (0 = No, 1 = ELL)	-10.73	3.72	-0.12	-2.88	0.004*
FRL (0 = No, 1 = FRL)	-6.80	1.96	-0.14	-3.46	0.001*
Gender (0 = M, 1= F)	5.04	1.88	0.11	2.68	0.008*
Block 2: Demographics + Fall 2011 PASS MC Score Model Fit: $F(5, 515) = 31.98$, $p < .001$, $R^2 = .24$, F Change $(1, 515) = 55.46$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-12.68	3.07	-0.17	-4.13	<0.001*
ELL (0 = No, 1 = ELL)	-5.63	3.61	-0.07	-1.56	0.119
FRL (0 = No, 1 = FRL)	-2.80	1.94	-0.06	-1.44	0.151
Gender (0 = M, 1= F)	4.48	1.79	0.10	2.50	0.013*
Fall PASS MC Scaled Score	0.07	0.01	0.33	7.45	<0.001*
Block 3: Demographics + Fall 2011 PASS MC Score + Phase Model Fit: $F(6, 514) = 30.93$, $p < .001$, $R^2 = .27$, F Change $(1, 514) = 19.81$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-13.49	3.02	-0.19	-4.47	<0.001*
ELL (0 = No, 1 = ELL)	-2.19	3.62	-0.03	-0.61	0.546
FRL (0 = No, 1 = FRL)	-1.56	1.93	-0.03	-0.81	0.420
Gender (0 = M, 1= F)	4.70	1.76	0.10	2.67	0.008*
Fall PASS MC Scaled Score	0.06	0.01	0.27	6.10	<0.001*
Phase (1= P1, 2 = P2)	-11.36	2.55	-0.19	-4.45	<0.001*

$p < 0.05$

Table 30. PASS Performance Task Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Middle School Cohort (N = 521): New Mexico

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	433	66.69	21.21	65.33	88	47.26	22.54	53.97	19.81	<0.000*	0.84	80
Not IEP	392	69.10	19.88	67.94	73	48.83	22.07	55.05	22.59	<0.000*	0.94	83
IEP	41	43.62	19.77	43.22	15	39.61	23.99	40.69	0.12	0.735	0.17	57
Not ELL	417	67.08	21.31	66.31	66	50.89	22.05	55.74	14.77	<0.000*	0.72	76
ELL	16	56.62	15.91	52.39	22	36.36	20.83	39.44	3.76	0.061	0.83	80
Not FRL	178	69.27	22.90	69.12	6	60.78	24.30	65.07	0.22	0.640	0.36	64
FRL	255	64.89	19.79	63.36	82	46.27	22.24	51.02	22.06	<0.000*	0.84	80
Male	201	61.22	22.56	59.82	41	49.93	23.24	56.82	0.64	0.425	0.43	67
Female	232	71.42	18.76	70.07	47	44.93	21.89	51.61	28.44	<0.000*	1.30	90

* $p < 0.05$

North Carolina Region: Results for Spring 2014 PASS Open-Ended and Performance Task

North Carolina
Spring 2014 PASS Open-Ended and Performance Task
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 PASS Open-Ended and Performance Task sections.

All

- Elementary Cohort Open-Ended: After controlling for the nearly substantively important advantage Phase 2 students demonstrated on the pretest (g = -0.23), Phase 1 students demonstrated a positive achievement advantage on the posttest over Phase 2 students, although the difference was not substantively important (g = 0.05).
- Middle School Performance Task: After controlling for the substantively important advantage Phase 2 students demonstrated on the pretest (g = -0.33), there was no substantively important difference on the posttest between the groups (g = -0.13).

ELL

- Elementary Cohort Open-Ended: While the Phase 2 ELL subgroup had a substantively important advantage on the pretest (g = -0.31), Phase 1 students outperformed Phase 2 on the posttest, although the effect size was not substantively important (g = 0.18).
- Elementary Cohort Performance Task: After controlling for the substantively important advantage Phase 2 ELL students had on the pretest (g = -0.31), there was no substantively important difference on the posttest between the groups (g = -0.07).
- Middle School Performance Task: After controlling for the substantively important advantage Phase 2 ELL students demonstrated on the pretest (g = -0.35), Phase 1 students outperformed Phase 2 students on the posttest, with a small but positive effect size (g = 0.10). It should be noted that the sample sizes for the ELL subgroup (Phase 1 n = 26; Phase 2 n = 28) were small.

Females

- Elementary Cohort Open-Ended: While the Phase 2 Female subgroup had a substantively important advantage on the pretest (g =-0.30), the effect size on the posttest (g = 0.04) favored Phase 1 students.
- After controlling for the substantively important advantage of Phase 2 Female students on the pretest, there was no substantively important difference on the posttest between the groups on
    - the Elementary Cohort Performance Task, or
    - the Middle School Cohort Open-Ended and Performance Task sections.

Economically Disadvantaged (FRL)

- Middle School Performance Task: After controlling for the substantively important advantage Phase 2 FRL students demonstrated on the pretest (g = -0.27), there was no substantively important difference on the posttest between the groups (g = -0.09)

IEP

- Elementary Cohort Open-Ended: After controlling for the slight Phase 2 baseline achievement advantage, the effect size for the IEP subgroup (*g* = 0.24) nearly reach the substantively important level.
- Middle School Cohort Open-Ended: After controlling for the substantively important advantage of Phase 2 IEP students on the pretest (*g* =-0.30), Phase 1 IEP students were able to demonstrate a very small, but positive effect size on the posttest (*g* =0.01).
- Elementary Cohort Performance Task: Even though Phase 2 students had an advantage on the pretest overall and for all subgroups, for the IEP subgroup, after controlling for pretest differences (*g* = -0.02), Phase 1 students outperformed Phase 2 students on the posttest, with the only substantively important posttest effect size (*g* = 0.28) across all groups.
- Middle School Performance Task: After controlling for the substantively important advantage Phase 2 IEP students demonstrated on the pretest (*g* = -0.41), Phase 1 students outperformed Phase 2 students on the posttest, with a nearly substantively meaningful effect size (*g* = 0.23).

Males

- Middle School Performance Task: After controlling for the substantively important advantage Phase 2 Male students demonstrated on the pretest (*g* = -0.30), there was no substantively important difference on the posttest between the groups (*g* = -0.03).

Spring 2014 PASS Open-Ended and Performance Task Results: North Carolina

A preliminary analysis was conducted on the Spring 2012 OE and PT sections of the PASS to determine baseline equivalence between Phase 1 and Phase 2 students in the elementary and middle school cohorts included the present analysis (see Table 31) as the PASS OE and PT sections were not administered until the end of the first posttest year, meaning there was no Fall 2011 baseline scores available. 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, 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).

Results indicated that for the elementary cohort aggregate scores (i.e., for all students combined), Phase 2 students had a statistically significantly higher mean Spring 2012 PT percent correct compared to Phase 1, with the effect size being substantially important according to What Work Clearinghouse (WWC) standards. For the middle school cohort aggregate scores, Phase 1 students had statistically significantly higher mean Spring 2012 OE and Spring 2012 PT percent correct scores, with the magnitude of the effect for Spring 2012 PT being substantially important.

Table 31. PASS OE and PT, Spring 2012 For Students Who Had a Spring 14 OE or PT Score, Treatment (Phase 1) and Control (Phase 2) Means Comparison: North Carolina

		Treatment (Phase 1)			Control (Phase 2)				
Section	Cohort	n	M	SD	n	M	SD	t	*g*
Open-Ended	Elementary	543	42.42	19.77	642	44.42	18.35	-1.8	-0.11
Performance Task	Elementary	604	50.92	18.41	677	55.68	17.45	-4.74*	-0.27
Open-Ended	Middle School	369	73.01	15.96	466	68.86	20.03	3.25*	0.23
Performance Task	Middle School	297	48.19	20.19	439	42.4	23.17	3.50*	0.26

* $p < 0.05$

Due to the fact that the PASS OE and PT were not administered until the end of the first posttest year, meaning there were no true baseline scores available, and due to substantively meaningful differences on the Spring 2012 scores, correlation analyses were conducted to examine the relationship between the Spring 2014 PASS OE and PT percent correct and (1) the Spring 2012 PASS OE and PT percent correct, as well as (2) the Fall 2011 PASS Multiple Choice (MC) scaled score results, to determine which scores would serve as the better baseline measure of achievement. The analyses revealed statistically significant, but low correlations among each of the measures of achievement (see Table 32). For the elementary cohort, the correlations for the available measures were both very similar, however, for the middle school cohort, the Fall 2011 PASS MC scaled scores had higher statistically significant correlations with the Spring 2014 PASS OE and PT.

Table 32. Correlations on the Percent Correct for Spring 2014 PASS OE and PT with Spring 2012 PASS OE and PT, and Fall 2011 PASS Multiple Choice: North Carolina

Spring 2014 PASS	Cohort	Fall 2011 PASS Multiple Choice	Spring 2012 Open-Ended	Spring 2012 Performance Task
Spring 2014 Open-Ended	Elementary	0.35*	0.32*	NA
	Middle School	0.46*	0.38*	NA
Spring 2014 Performance Task	Elementary	0.33*	NA	0.35*
	Middle School	0.40*	NA	0.24*

* $p < 0.05$

To determine baseline equivalence on the Fall 2011 PASS MC between North Carolina Phase 1 and Phase 2 students included the present analyses, a series of independent *t*-tests were 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 (see Table 33). For the elementary OE cohort in the aggregate (i.e., the "All" group), Phase 2 students demonstrated a statistically significant advantage over their Phase 1 counterparts in their baseline achievement levels ($t(1328) = -4.17$, $p < 0.001$, $g = -0.23$, PR = 41), but the effect size linked to this advantage did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$). Consistent with this overall difference in performance, statistically significant, but not substantively important advantages were observed to favor four subgroups of Phase 2 students in the elementary cohort: namely, Not IEP, Not ELL, FRL, and Males. Among elementary cohort students who were female, however, the advantage seen for those in Phase 2 proved not only to be statistically significant, but also substantively important ($g = -0.30$). Additionally, ELL Phase 2 students had an advantage over Phase 1 students that was not statistically significant, but was substantively important ($g = -0.31$).

For students in the middle school OE cohort, similar advantages in baseline achievement levels were observed for Phase 2 over Phase 1 students. As with the elementary OE cohort, a statistically significant difference in aggregate performance (i.e., the "All" group) for Phase 2 students was observed ($t(969) = -3.64$, $p < 0.001$, $g = -0.24$, PR = 41), but the effect size linked to this difference did not meet WWC criteria for substantive importance. Statistically significant, but not substantively important advantages in baseline performance were also observed for four subgroups of Phase 2 middle school cohort students: Not IEP, Not ELL, FRL, and Males. Among middle school cohort students who were female, however, the advantage seen for those in Phase 2 proved not only to be statistically significant, but also substantively important ($g = -0.27$). Additionally, IEP Phase 2 students had an advantage over Phase 1 students that was not statistically significant, but was substantively important ($g = -0.30$).

With respect to the elementary PT cohort in the aggregate (i.e., the "All" group), Phase 2 students demonstrated a statistically significant advantage over their Phase 1 counterparts in their baseline achievement levels ($t(1326) = -4.18$, $p < 0.001$, $g = -0.23$, PR = 41), but the effect size linked to this advantage did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$). Consistent with this overall difference in performance, statistically significant, but not substantively important advantages were observed to favor four subgroups of Phase 2 students in the elementary cohort: namely, Not IEP, Not ELL, FRL, and Males. Among elementary cohort students who were female, however, the advantage seen for those in Phase 2 proved not only to be statistically significant, but also substantively important ($g = 0.29$). Additionally, ELL Phase 2 students had an advantage over Phase 1 students that was not statistically significant, but was substantively important ($g = -0.31$).

With respect to students in the middle school PT cohort, a statistically significant difference in aggregate performance (i.e., the “All” group) favoring Phase 2 students was observed ($t(885) = -4.82$, $p < 0.001$, $g = -0.33$, PR = 37), and linked to an effect size that exceeded the WWC criteria for substantive importance. Statistically significant and substantively important advantages in baseline performance were also observed for five subgroups of Phase 2 middle school cohort students: Not IEP, Not ELL, FRL, Males, and Females. Among middle school cohort students who were classified as IEP or ELL, however, the advantage seen for those in Phase 2 proved not to be statistically significant, but substantively important ($g = -0.41$ and $g = -0.35$ respectively).

While neither the Spring 2012 PASS OE and PT nor the Fall 2011 PASS Multiple Choice provided complete baseline equivalence between Phase 1 and Phase 2 students for all students combined, the Fall 2011 PASS Multiple Choice was administered as a true baseline assessment vs. the Spring 2012 PASS OE and PT, which was not administered until the end of the first posttest year. Therefore, due to its stronger relationship to the Spring 2014 PASS OE and PT outcomes, and because it was a true baseline measure, the Fall 2011 PASS Multiple Choice scaled score was chosen as the covariate (i.e., pretest measure) for both the elementary and middle school cohorts.

Table 33. Fall 2011 PASS Multiple Choice for Students Who Had Spring 2014 OE or PT Scores, Treatment (Phase 1) and Control (Phase 2) Means Comparison: North Carolina

	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort - Open-Ended									
All	622	304.4	95.64	708	325.97	92.84	-4.17*	-0.23	41
Not IEP	558	310.01	95.28	649	332.27	90.65	-4.15*	-0.24	41
IEP	64	255.44	84.76	59	256.73	89.05	-0.08	-0.01	49
Not ELL	534	312.4	95.24	666	328.82	93.46	-3.00*	-0.17	43
ELL	88	255.88	83.34	42	280.81	68.76	-1.68	-0.31	38
Not FRL	272	341.67	86.23	397	350.41	92.53	-1.23	-0.1	46
FRL	350	275.44	92.62	311	294.79	83.54	-2.81*	-0.22	41
Male	311	310.02	100.76	366	326.35	95.4	-2.16*	-0.17	43
Female	311	298.77	90.05	342	325.57	90.15	-3.80*	-0.3	38
	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Middle School Cohort - Open-Ended									
All	389	358.11	106.92	582	382.8	101.09	-3.64*	-0.24	41
Not IEP	348	368.14	103.49	533	390	97.19	-3.18*	-0.22	41
IEP	41	273.02	98.42	49	304.49	110.32	-1.42	-0.3	38
Not ELL	361	365.13	105.92	551	388.3	98.97	-3.36*	-0.23	41
ELL	28	267.57	74.47	31	285	88.86	-0.81	-0.21	42
Not FRL	160	411.85	95.44	309	418.31	87.45	-0.74	-0.07	47
FRL	229	320.56	98.29	273	342.6	100.57	-2.47*	-0.22	41
Male	193	365.87	111.39	287	387.96	101.15	-2.25*	-0.21	42
Female	196	350.47	102.05	295	377.78	100.95	-2.92*	-0.27	39
	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Elementary Cohort - Performance Task									
All	626	304.4	95.51	702	326.02	92.85	-4.18*	-0.23	41
Not IEP	561	310.12	95.16	643	332.38	90.64	-4.15*	-0.24	41
IEP	65	255.02	84.16	59	256.73	89.05	-0.11	-0.02	49
Not ELL	538	312.34	95.09	660	328.9	93.48	-3.03*	-0.18	43
ELL	88	255.88	83.34	42	280.81	68.76	-1.68	-0.31	38
Not FRL	275	341.77	85.98	393	350.39	92.48	-1.22	-0.1	46
FRL	351	275.13	92.44	309	295.02	83.75	-2.88*	-0.22	41
Male	313	310.02	100.76	363	326.77	95.35	-2.22*	-0.17	43
Female	313	298.78	89.77	339	325.22	90.23	-3.75*	-0.29	38
	Treatment (Phase 1)			Control (Phase 2)					
Group	*n*	*M*	*SD*	*n*	*M*	*SD*	*t*	*g*	PR
Middle School Cohort - Performance Task									
All	339	350.01	110.26	548	384.46	99	-4.82*	-0.33	37
Not IEP	296	362.48	106.31	502	391.62	95.4	-3.99*	-0.29	39
IEP	43	264.16	99.04	46	306.35	104.85	-1.95	-0.41	34
Not ELL	313	356.81	109.97	520	389.21	97.63	-4.42*	-0.32	38
ELL	26	268.23	76.86	28	296.32	82.78	-1.29	-0.35	36
Not FRL	129	402.71	105.08	292	419.13	84.61	-1.7	-0.18	43
FRL	210	317.64	100.65	256	344.91	99.54	-2.93*	-0.27	39
Male	166	360.18	112.24	267	391.29	98.74	-3.02*	-0.3	38
Female	173	340.25	107.74	281	377.97	98.98	-3.81*	-0.37	36

* $p < 0.05$

Employing the Fall 2011 PASS MC data as covariates to statistically adjust the outcomes for baseline differences in achievement, preliminary analyses were conducted on Spring 2014 PASS OE and PT percent correct scores to determine any differences between Phase 1 and Phase 2 North Carolina students in the elementary and middle school cohorts. As these analyses were exploratory in nature, no corrections were made for multiple comparisons. As noted earlier, for both the elementary and middle school cohorts, Phase 1 and Phase 2 students were not equivalent on the baseline measure for the OE and PT, with Phase 2 students having an advantage both overall (for the middle school PT), and for several subgroups on both the OE and PT. Due to these baseline differences, there were violations of the ANCOVA assumption of equal variances. Therefore, results for these particular groups should be interpreted with the Phase 2 advantage and the statistical issues in mind.

Elementary and Middle School Cohorts PASS Open-Ended Analyses: North Carolina

With respect to the 1,330 elementary cohort students in Phase 1 (*n* = 622) and Phase 2 (*n* = 708) schools and the 971 middle school cohort students in Phase 1 (*n* = 389) and Phase 2 (*n* =582) schools, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their percent correct score on the Spring 2014 OE section of the PASS assessment. In addition to these regressions, a second set of ANCOVA analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (*g*), was also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their Special Education (IEP) status, English Language Learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender.

Elementary Cohort Spring 2014 PASS Open-Ended Results: North Carolina

Among the 1,330 elementary cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS Multiple Choice scaled scores (Block 3) explained 16% of the total variance (R^2) in students' Spring 2014 OE scores (see Table 34). Adding Phase to the model did not increase the variance explained, and Phase did not have a statistically significant impact on the Spring 2014 OE percent correct (β = -0.03, t = -1.06, p = 0.290).

While the ANCOVA adjusted means presented in Table 35 can be seen to trend higher for Phase 1 students (*n* = 622, Adjusted Mean = 67.06) than for Phase 2 students (*n* = 708, Adjusted Mean = 65.97) overall (i.e., the "All" group) (*F* (1, 1323) = 1.12, *p* = 0.290, *g* = 0.05, *PR* = 52), the difference was not statistically significant, nor was the magnitude of the effect size (*g* = 0.05) considered to be substantively important. No subgroup comparisons indicated either statistically significant or substantively important differences between the adjusted means for Phase 1 and Phase 2. However, Phase 1 students demonstrated a positive achievement advantage on the posttest over Phase 2 students both overall and for each subgroup after controlling for the Phase 2 baseline achievement advantage for all groups. In addition, the effect size for the IEP subgroup (*g* = 0.24) nearly reached the substantively important level. Furthermore, while the Phase 2 ELL (*g* = -0.31) and Female (*g* =-0.30) subgroups had substantively important advantages on the pretest, both subgroup effect sizes on the posttest favored Phase 1 students, although neither was substantively important.

Table 34. PASS Open-Ended Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Elementary Cohort (N = 1,330): North Carolina

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1325) = 29.96$, $p < .001$, $R^2 = .08$, *F* Change (4, 1325) = 29.96, $p < .001$)					
IEP (0 = No, 1 = IEP)	-14.17	1.84	-0.20	-7.70	<0.001*
ELL (0 = No, 1 = ELL)	-1.89	1.87	-0.03	-1.01	0.312
FRL (0 = No, 1 = FRL)	-6.47	1.11	-0.16	-5.82	<0.001*
Gender (0 = M, 1= F)	3.23	1.06	0.08	3.04	0.002*
Block 2: Demographics + Fall 2011 PASS Score Model Fit: $F(5, 1324) = 49.89$, $p < .001$, $R^2 = .16$, *F* Change (1, 1324) = 118.95, $p < .001$)					
IEP (0 = No, 1 = IEP)	-10.17	1.80	-0.15	-5.65	<0.001*
ELL (0 = No, 1 = ELL)	0.16	1.80	0.00	0.09	0.932
FRL (0 = No, 1 = FRL)	-3.00	1.11	-0.08	-2.70	0.007*
Gender (0 = M, 1= F)	3.75	1.02	0.09	3.68	<0.001*
Fall 2011 PASS Scaled Score	0.06	0.01	0.30	10.91	<0.001*
Block 3: Demographics + Fall 2011 PASS Score + Phase Model Fit: $F(6, 1323) = 41.77$, $p < .001$, $R^2 = .16$, *F* Change (1, 1323) = 1.12, $p = .290$)					
IEP (0 = No, 1 = IEP)	-10.21	1.80	-0.15	-5.67	<0.001*
ELL (0 = No, 1 = ELL)	-0.04	1.81	-0.00	-0.02	0.982
FRL (0 = No, 1 = FRL)	-3.08	1.11	-0.08	-2.76	0.006*
Gender (0 = M, 1= F)	3.74	1.02	0.09	3.67	<0.001*
Fall 2011 PASS Scaled Score	0.06	0.01	0.30	10.95	<0.001*
Phase (1= P1, 2 = P2)	-1.09	1.03	-0.03	-1.06	0.290

$p < 0.05$

Table 35. PASS Open-Ended Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Elementary Cohort (N = 1,330): North Carolina

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	622	66.05	21.17	67.06	708	66.86	19.18	65.97	1.12	0.290	0.05	52
Not IEP	558	67.26	20.51	68.25	649	68.41	18.11	67.57	0.42	0.518	0.04	51
IEP	64	55.47	23.95	55.35	59	49.72	22.21	49.84	1.73	0.191	0.24	59
Not ELL	534	66.57	20.84	67.39	666	67.19	19.13	66.53	0.65	0.419	0.04	52
ELL	88	62.88	22.98	63.74	42	61.51	19.30	59.71	0.99	0.322	0.18	57
Not FRL	272	70.40	20.23	71.03	397	69.86	18.75	69.43	1.29	0.256	0.08	53
FRL	350	62.67	21.30	63.14	311	63.02	19.05	62.49	0.19	0.666	0.03	51
Male	311	64.42	22.21	65.37	366	64.75	19.84	63.94	0.89	0.346	0.07	53
Female	311	67.68	19.99	68.83	342	69.10	18.20	68.06	0.29	0.588	0.04	52

* $p < 0.05$

Middle School Cohort Spring 2014 PASS Open-Ended Results: North Carolina

Among the 971 middle school cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS MC scaled scores (Block 3) explained 28% of the total variance (R^2) in students' 2014 Spring OE scores. Adding Phase to the model only increased the variance explained by 1%, and Phase did not have a statistically significant impact on the Spring 2014 OE percent correct, (β = 0.03, t = 1.12, p = 0.263) (see Table 36).

The ANCOVA adjusted means presented in Table 37 indicate no statistically significant difference between Phase 1 students (n = 389, Adjusted Mean = 83.78) and Phase 2 students (n = 582, Adjusted Mean = 84.75) overall (i.e., the "All" group) (F (1, 964) = 1.25, p = 0.263, g = -0.06, PR = 47). Only one subgroup comparison (ELL) was statistically significant (F (1, 53) = 4.06, p = 0.049, g = -0.53, PR = 30), favoring Phase 2 students, producing an effect size that was substantively important (with Phase 2 ELL students also having an advantage (g = -0.21) on the pretest). However, after controlling for the substantively important advantage of Phase 2 IEP (g =-0.30) and Female (g =-0.27) students on the pretest, Phase 1 IEP students were able to demonstrate a very small, but positive effect size (g =0.01) on the posttest, and the posttest effect size for Females was small and not substantively important (g = -0.06).

Table 36. PASS Open-Ended Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Middle School Cohort (N = 971): North Carolina

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 966) = 48.65$, $p < .001$, $R^2 = .17$, F Change $(4, 966) = 48.65$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-11.64	1.61	-0.22	-7.24	<0.001*
ELL (0 = No, 1 = ELL)	-8.13	1.96	-0.13	-4.15	<0.001*
FRL (0 = No, 1 = FRL)	-7.02	0.92	-0.23	-7.66	<0.001*
Gender (0 = M, 1= F)	2.70	0.90	0.09	3.00	0.003*
Block 2: Demographics + Fall 2011 PASS MC Score Model Fit: $F(5, 965) = 72.85$, $p < .001$, $R^2 = .27$, F Change $(1, 965) = 141.38$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-8.12	1.53	-0.15	-5.30	<0.001*
ELL (0 = No, 1 = ELL)	-5.19	1.85	-0.08	-2.81	0.005*
FRL (0 = No, 1 = FRL)	-3.03	0.92	-0.10	-3.29	0.001*
Gender (0 = M, 1= F)	3.54	0.84	0.12	4.20	<0.001*
Fall 2011 PASS MC SS	0.05	0.01	0.37	11.89	<0.001*
Block 3: Demographics + Fall 2011 PASS MC Score + Phase Model Fit: $F(6, 964) = 60.93$, $p < .001$, $R^2 = .28$, F Change $(1, 964) = 1.25$, $p = .263$)					
IEP (0 = No, 1 = IEP)	-8.12	1.53	-0.15	-5.30	<0.001*
ELL (0 = No, 1 = ELL)	-5.19	1.85	-0.08	-2.81	0.005*
FRL (0 = No, 1 = FRL)	-2.95	0.92	-0.10	-3.19	0.001*
Gender (0 = M, 1= F)	3.54	0.84	0.12	4.19	<0.001*
Fall 2011 PASS MC SS	0.05	0.01	0.37	11.78	<0.001*
Phase (1= P1, 2 = P2)	0.97	0.86	0.03	1.12	0.263

$p < 0.05$

Table 37. PASS Open-Ended Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Middle School Cohort (N = 971): North Carolina

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	389	82.61	16.07	83.78	582	85.53	14.61	84.75	1.25	0.263	-0.06	47
Not IEP	348	84.10	15.02	85.06	533	86.88	12.92	86.26	2.00	0.157	-0.09	47
IEP	41	69.92	19.10	70.50	49	70.88	22.26	70.39	0.00	0.979	0.01	50
Not ELL	361	83.95	14.98	84.98	551	86.06	14.41	85.39	0.22	0.642	-0.03	49
ELL	28	65.24	19.57	66.03	31	76.13	15.28	75.42	4.06	0.049*	-0.53	30
Not FRL	160	87.54	13.48	87.74	309	89.54	10.98	89.43	2.57	0.110	-0.14	44
FRL	229	79.16	16.84	79.99	273	81.00	16.76	80.31	0.06	0.812	-0.02	49
Male	193	80.90	15.74	82.07	287	83.93	15.84	83.14	0.71	0.400	-0.07	47
Female	196	84.29	16.25	85.42	295	87.10	13.15	86.35	0.63	0.428	-0.06	47

* $p < 0.05$

Elementary and Middle School Cohorts PASS Performance Task Analyses: North Carolina

With respect to the 1,328 elementary cohort students in Phase 1 (*n* = 626) and Phase 2 (*n* = 702) schools and the 887 middle school cohort students in Phase 1 (*n* = 339) and Phase 2 (*n* =548) schools, hierarchical or "block entry" multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their percent correct score on the Spring 2014 PT section of the PASS assessment. In addition to these regressions, a second set of ANCOVA analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (*g*), was also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their Special Education (IEP) status, English Language Learner (ELL) status, Economically Disadvantaged (FRL) status, and Gender.

Elementary Cohort Spring 2014 PASS Performance Task Results: North Carolina

Among the 1,328 elementary cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS Multiple Choice scaled scores (Block 3) explained 16% of the total variance (R^2) in students' Spring 2014 PT scores (see Table 38). While adding Phase to the model only increased the variance explained by 1%, there was a statistically significant difference in the Spring 2014 PT percent correct favoring Phase 2 students (β = 0.05, t = 2.08, p = 0.038).

The ANCOVA adjusted means presented in Table 39 can be seen to be statistically significantly higher for Phase 2 students (n = 702, Adjusted Mean = 67.61) than for Phase 1 students (n = 626, Adjusted Mean = 65.97) overall (i.e., the "All" group) (F (1, 1321) = 4.33, p = 0.038, g = -0.11, PR = 46), indicating that the average Phase 1 student scored at the 46th percentile of the control group. However, the effect size was not considered to be substantively important according to WWC standards. Consistent with these overall outcomes, three subgroup analyses (Not IEP, Not ELL, and Females) were statistically significant and all favored Phase 2 elementary cohort students, as indicated by the negatively signed effect sizes, but none were substantively meaningful (see Table 39). Meanwhile, even though Phase 2 students had an advantage on the pretest overall and for all subgroups, for the IEP subgroup, after controlling for pretest differences (g = -0.02), Phase 1 students outperformed Phase 2 students on the posttest, with the only substantively important posttest effect size across all groups (g = 0.28). In addition, after controlling for the substantively important advantage Phase 2 ELL students (g = -0.31) and females (g = -0.29) had on the pretest, there were no substantively important differences on the posttest between the groups (g = -0.07 and g = -0.17 respectively).

Table 38. PASS Performance Task Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Elementary Cohort (N = 1,328): North Carolina

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 1323) = 35.09$, $p < .001$, $R^2 = .10$, *F* Change (4, 1323) = 35.09, $p < .001$)					
IEP (0 = No, 1 = IEP)	-10.72	1.38	-0.20	-7.76	<0.001*
ELL (0 = No, 1 = ELL)	-3.33	1.41	-0.07	-2.37	0.018*
FRL (0 = No, 1 = FRL)	-5.55	0.84	-0.18	-6.61	<0.001*
Gender (0 = M, 1= F)	2.12	0.80	0.07	2.65	0.008*
Block 2: Demographics + Fall PASS MC Score Model Fit: $F(5, 1322) = 47.62$, $p < .001$, $R^2 = .15$, *F* Change (1, 1322) = 88.46, $p < .001$)					
IEP (0 = No, 1 = IEP)	-8.08	1.37	-0.15	-5.91	<0.001*
ELL (0 = No, 1 = ELL)	-1.99	1.37	-0.04	-1.45	0.148
FRL (0 = No, 1 = FRL)	-3.27	0.85	-0.11	-3.86	<0.001*
Gender (0 = M, 1= F)	2.48	0.78	0.08	3.20	0.001*
Fall PASS MC Scaled Score	0.04	0.00	0.26	9.41	<0.001*
Block 3: Demographics + Fall PASS MC Score + Phase Model Fit: $F(6, 1321) = 40.50$, $p < .001$, $R^2 = .16$, *F* Change (1, 1321) = 4.33, $p = .038$)					
IEP (0 = No, 1 = IEP)	-8.02	1.37	-0.15	-5.87	<0.001*
ELL (0 = No, 1 = ELL)	-1.70	1.38	-0.03	-1.23	0.217
FRL (0 = No, 1 = FRL)	-3.16	0.85	-0.10	-3.73	<0.001*
Gender (0 = M, 1= F)	2.51	0.78	0.08	3.23	0.001*
Fall PASS MC Scaled Score	0.04	0.00	0.26	9.26	<0.001*
Phase (1= P1, 2 = P2)	1.64	0.79	0.05	2.08	0.038*

* $p < 0.05$

Table 39. PASS Performance Task Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Elementary Cohort (N = 1,328): North Carolina

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	626	65.17	15.36	65.97	702	68.32	15.07	67.61	4.33	0.038*	-0.11	46
Not IEP	561	65.92	15.02	66.67	643	69.64	14.01	68.99	8.26	0.004*	-0.16	44
IEP	65	58.64	16.80	58.78	59	54.04	18.55	53.88	2.57	0.111	0.28	61
Not ELL	538	65.83	15.12	66.43	660	68.65	14.97	68.17	4.62	0.032*	-0.12	45
ELL	88	61.10	16.29	61.38	42	63.17	15.81	62.56	0.15	0.697	-0.07	47
Not FRL	275	68.92	14.43	69.36	393	70.86	14.00	70.55	1.33	0.249	-0.08	47
FRL	351	62.23	15.45	62.59	309	65.11	15.77	64.69	3.12	0.078	-0.13	45
Male	313	64.52	15.17	65.16	363	66.50	15.60	65.96	0.52	0.471	-0.05	48
Female	313	65.81	15.55	66.79	339	70.28	14.25	69.37	5.36	0.021*	-0.17	43

* $p < 0.05$

Middle School Cohort Spring 2014 PASS Performance Task Results: North Carolina

Among the 887 middle school cohort students across the region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS MC scaled scores (Block 3) explained 19% of the total variance (R^2) in students' 2014 Spring PT scores (see Table 40). While adding Phase to the model did not increase the variance explained, Phase did have a statistically significant impact on the Spring 2014 PT percent correct, favoring Phase 2 students ($\beta = 0.06$, $t = 2.06$, $p = 0.040$).

While Phase 2 students had substantively important advantages on the pretest for all groups except not Economically Disadvantaged, none of the effect sizes for any groups on the posttest were substantively important. In addition, the Phase 1 IEP ($g = 0.23$) and ELL subgroups ($g = 0.10$) demonstrated positive effect sizes on the posttest, with the effect size for the IEP subgroup nearly reaching the substantively important level. Therefore, while the ANCOVA adjusted means presented in Table 41 can be seen to be statistically significantly higher for Phase 2 students ($n = 548$, Adjusted Mean = 53.67) than for Phase 1 students ($n = 339$, Adjusted Mean = 50.55) overall (i.e., the "All" group) (F (1, 880) = 4.25, $p = 0.040$, $g = -0.13$, $PR = 45$), indicating that the average Phase 1 student scored at the 45th percentile of the control group, the Phase 2 group had a substantively meaningful advantage on the pretest. It should also be noted that the sample sizes for the ELL subgroup (Phase 1 $n = 26$; Phase 2 $n = 28$) were small, and therefore may not be representative.

Table 40. PASS Performance Task Questions, Spring 2014: Mean Hierarchical Multiple Regression Summary for the Middle School Cohort (N = 887): North Carolina

Source	*B*	*S.E.B.*	*β*	*t*	*p*
Block 1: Demographics Model Fit: $F(4, 882) = 23.82$, $p < .001$, $R^2 = .10$, F Change $(4, 882) = 23.82$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-12.67	2.62	-0.16	-4.83	<0.001*
ELL (0 = No, 1 = ELL)	-8.30	3.31	-0.08	-2.50	0.013*
FRL (0 = No, 1 = FRL)	-9.40	1.56	-0.20	-6.03	<0.001*
Gender (0 = M, 1= F)	2.46	1.53	0.05	1.61	0.108
Block 2: Demographics + Fall 2011 PASS MC Score Model Fit: $F(5, 881) = 39.93$, $p < .001$, $R^2 = .19$, F Change $(1, 881) = 94.27$, $p < .001$)					
IEP (0 = No, 1 = IEP)	-7.08	2.56	-0.09	-2.77	0.006*
ELL (0 = No, 1 = ELL)	-4.98	3.17	-0.05	-1.57	0.117
FRL (0 = No, 1 = FRL)	-3.95	1.59	-0.08	-2.49	0.013*
Gender (0 = M, 1= F)	4.08	1.46	0.09	2.79	0.005*
Fall PASS MC Scaled Score	0.08	0.01	0.33	9.71	<0.001*
Block 3: Demographics + Fall 2011 PASS MC Score + Phase Model Fit: $F(6, 880) = 34.10$, $p < .001$, $R^2 = .19$, F Change $(1, 880) = 4.25$, $p = .040$)					
IEP (0 = No, 1 = IEP)	-6.96	2.55	-0.09	-2.73	0.007*
ELL (0 = No, 1 = ELL)	-4.96	3.16	-0.05	-1.57	0.117
FRL (0 = No, 1 = FRL)	-3.65	1.59	-0.08	-2.29	0.022*
Gender (0 = M, 1= F)	4.06	1.46	0.09	2.78	0.006*
Fall PASS MC Scaled Score	0.07	0.01	0.33	9.46	<0.001*
Phase (1= P1, 2 = P2)	3.12	1.51	0.06	2.06	0.040*

$p < 0.05$

Table 41. PASS Performance Task Questions, Spring 2014: Mean Comparison of Phase 1 (Treatment) and Phase 2 (Control) Middle School Cohort (N = 887): North Carolina

	Treatment (Phase 1)				Control (Phase 2)							
Area	n	M	SD	Adj. M	n	M	SD	Adj. M	F	*p*	*g*	PR
All	339	48.36	24.57	50.55	548	55.02	22.92	53.67	4.25	0.040*	-0.13	45
Not IEP	296	49.80	24.76	51.58	502	56.71	22.57	55.66	6.32	0.012*	-0.17	43
IEP	43	38.44	20.95	39.78	46	36.57	18.26	35.32	1.26	0.266	0.23	59
Not ELL	313	49.22	24.75	51.23	520	55.97	22.83	54.77	5.06	0.025*	-0.15	44
ELL	26	38.01	19.91	38.70	28	37.40	16.73	36.76	0.14	0.713	0.10	54
Not FRL	129	54.40	26.66	55.24	292	60.17	22.90	59.80	3.47	0.063	-0.19	43
FRL	210	44.65	22.47	46.02	256	49.15	21.52	48.03	1.13	0.289	-0.09	46
Male	166	48.09	24.35	50.28	267	52.44	22.96	51.07	0.13	0.718	-0.03	49
Female	173	48.62	24.85	51.02	281	57.48	22.65	56.01	5.62	0.018*	-0.21	42

* $p < 0.05$.

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