



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 3:
PASS Assessments Multiple Choice

Brenda McSparrin-Gallagher, Ed. D.

Yun Tang, B.A.

Brian Niemeier, Ph.D.

Todd Zoblotsky, Ph.D.

The University of Memphis

7/15/2015

Acknowledgments

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

CREP Project Staff:

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

Introduction

There were a total of 29 multiple choice questions on both the Fall 2011 and Spring 2014 forms of the PASS (PASS-MC) addressing five broad science content standard categories for the elementary cohort and six broad science content standard categories for the middle school cohort. Only students who answered at least one multiple choice achievement question at both time points were included in the analyses for each respective area of analysis.

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) and subgroup.

All Regions: Results for Spring 2014 PASS Multiple Choice

All Regions Spring 2014 PASS Multiple Choice Key Findings for Phase 1

For all students combined (the “All” group) and the specified subgroups across all three regions, the following outcomes favoring Phase 1 students were found on the Spring 2014 PASS multiple choice section.

Not Economically Disadvantaged (Not FRL)

- Elementary Cohort: Phase 1 students scored statistically significantly higher than Phase 2 students, but the difference was not substantively important ($g = 0.10$).

Fall 2011 to Spring 2014 PASS Results: All Regions

PASS-Multiple Choice: All Regions

Table 1 shows the final sample sizes employed in the elementary cohort analyses (currently 5th graders in 2013-2014) once students missing data on all 29 PASS-MC questions at either time point were excluded.

Table 1. PASS-MC, Spring 2014: Samples for the PASS-MC Analyses for the Elementary Cohort: All Regions

Sample	Phase 1	Phase 2
Students available for the PASS-MC achievement analysis	2,338	1,785

Table 2 shows the final sample sizes employed in the middle school cohort analyses (currently 8th graders in 2013-2014) once students missing all 29 PASS-MC questions at either time point were excluded.

Table 2. PASS-MC, Spring 2014: Samples for the PASS-MC Analyses for the Middle School Cohort: All Regions

Sample	Phase 1	Phase 2
Students available for the PASS-MC achievement analysis	1,036	1,132

To determine baseline equivalence on the Fall 2011 PASS-MC between Phase 1 and Phase 2 students included the present analysis, 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. In addition, an effect size was also calculated as a measure of baseline equivalence.

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

With respect to the elementary cohort (Table 3), students in the aggregate (the “All” group) did not demonstrate a statistically significant difference by Phase in their baseline achievement levels ($t(4121) = -0.75, p = .45, g = -0.02, PR = 49$). At the same time, ELL students were the only subgroup that appeared to have a statistically significant difference in baseline achievement, with Phase 2 ELL students outperforming their Phase 1 counterparts, although based on the effect size (*g*), not to a substantively meaningful degree ($t(926.9) = -2.36, p = .02, g = -0.15, PR = 44$). Overall, there were no substantively important effect size differences for the elementary cohort, meaning there was baseline equivalence for all groups.

Table 3. Baseline Comparison of Fall 2011 PASS-MC Scaled Scores for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 4,123): All Regions

Group	Treatment (Phase 1)			Control (Phase 2)			<i>t</i>	<i>g</i>	PR
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>			
Elementary Cohort									
All	2,338	312.02	101.33	1,785	314.39	98.11	-0.75	-0.02	49
Not IEP	2,129	316.44	101.05	1,633	319.37	96.56	-0.91	-0.03	49
IEP	209	267.05	93.13	152	260.83	98.96	0.61	0.06	53
Not ELL	1,801	326.40	99.67	1,367	325.72	99.28	0.19	0.01	50
ELL	537	263.82	91.57	418	277.31	84.24	-2.36*	-0.15	44
Not FRL	922	353.86	96.65	725	350.93	97.68	0.61	0.03	51
FRL	1,416	284.79	94.85	1,060	289.39	90.29	-1.23	-0.05	48
Male	1,181	313.47	104.97	898	316.78	99.30	-0.73	-0.03	49
Female	1,157	310.55	97.49	887	311.97	96.89	-0.33	-0.01	49

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

* $p < .05$.

Likewise, with respect to students in the middle school cohort (Table 4), there was no statistically significant difference in baseline achievement by Phase ($t(2166) = 1.17, p = .24, g = 0.05, PR = 52$) in the aggregate. When the outcomes for FRL students were compared by Phase, there was a statistically significant difference in Fall 2011 PASS scores that favored Phase 1, but the effect size linked to the comparison did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$) ($t(1223.2) = 3.62, p < .01, g = 0.20, PR = 58$). On the other hand, there was a statistically significant difference in Fall 2011 PASS scores for ELL students, and the effect size associated with the difference met the WWC threshold for substantive importance, favoring Phase 1 students ($t(181) = 3.30, p < .01, g = 0.49, PR = 69$). Therefore, the outcome for the ELL subgroup comparison for the middle school cohort should be interpreted in light of the substantively important difference in baseline achievement between Phase 1 and Phase 2 students.

Employing these Fall 2011 data as covariates to statistically adjust the outcomes for baseline differences in achievement, preliminary analyses were conducted on Spring 2014 PASS-MC scaled scores to determine differences between Phase 1 and Phase 2 students in the elementary and middle school cohorts, with each student's scaled score on the Spring 2014 PASS-MC used as the outcome measure. As these analyses were exploratory in nature, no corrections were made for multiple comparisons. Additionally, some violations in the ANCOVA assumptions were observed for subgroup comparisons.

Table 4. Baseline Comparison of Fall 2011 PASS-MC Scaled Scores for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 2,168): All Regions

Group	Treatment (Phase 1)			Control (Phase 2)			t	g	PR
	n	M	SD	n	M	SD			
Middle School Cohort									
All	1,036	364.51	102.66	1,132	359.10	112.40	1.17	0.05	52
Not IEP	925	374.38	98.92	1,018	368.38	108.62	1.27	0.06	52
IEP	111	282.22	96.57	114	276.25	112.07	0.43	0.06	52
Not ELL	953	370.93	101.79	1,032	369.86	108.43	0.23	0.01	50
ELL	83	290.70	82.07	100	248.08	90.75	3.30*	0.49	69
Not FRL	392	406.28	95.31	518	408.26	92.72	-0.32	-0.02	49
FRL	644	339.08	98.64	614	317.63	110.88	3.62*	0.20	58
Male	505	361.17	105.54	570	360.51	116.65	0.10	0.01	50
Female	531	367.68	99.85	562	357.67	108.00	1.59	0.10	54

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

* $p < .05$.

Elementary and Middle School Cohort PASS-Multiple Choice Analyses: All Regions

With respect to the cohort of 4,123 elementary students in Phase 1 ($n = 2,338$) and Phase 2 ($n = 1,785$) schools and the cohort of 2,168 middle school students in Phase 1 ($n = 1,036$) and Phase 2 ($n = 1,132$) schools, hierarchical or “block entry” multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their performance on the Spring 2014 PASS-MC assessment scaled score (see Table 5 and Table 7). In addition to these regressions, a second set of analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (*g*) were also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their IEP status, ELL status, FRL status, and Gender (see Table 6 and Table 8).

Elementary Cohort PASS-Multiple Choice Spring 2014 Results: All Regions

For the 4,123 students across the three regions in the elementary cohort, 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 PASS-MC scores (see Table 5). The addition of the student’s Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of Spring 2014 PASS-MC achievement ($\beta = 0.01$, $t = 0.39$, $p = .698$).

While the overall (i.e., the “All” group) ANCOVA adjusted mean presented in Table 6 was higher for Phase 1 students ($n = 2,338$, Adjusted Mean = 435.80) compared to Phase 2 students ($n = 1,785$, Adjusted Mean = 434.88), it also fell short of being statistically significant ($F(1, 4116) = 0.15$, $p = 0.698$, $g = 0.01$, $PR = 50$), and the effect size ($g = 0.01$) was not substantively important. Consistent with these overall outcomes, five subgroup ANCOVA analyses (IEP, Not ELL, ELL, Not FRL, and Male) were linked to positively signed effects and favored Phase 1 students in the elementary cohort (see Table 6). Nevertheless, only one of the comparisons indicated a statistically significant difference between the

adjusted means: specifically, the comparison favoring the 922 Phase 1 students over the 725 Phase 2 students who were Not FRL ($F(1, 1641) = 5.14, p = 0.024, g = 0.10, PR = 54$), with an effect size ($g = 0.10$) that was not substantively important, and indicated that the average Phase 1 student scored at the 54th percentile of the control group. Additionally, for the ELL subgroup, Phase 2 statistically significantly outperformed Phase 1 at baseline, but on the posttest Phase 1 had a higher adjusted mean score that fell short of being statistically significant or substantively important. Overall, none of the effect sizes for the ANCOVA analyses were large enough to be substantively important, ranging from a low of -0.03 (FRL and Female) to a high of 0.19 (IEP).

Table 5. PASS-MC, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students (N = 4,123): All Regions

Source	B	S.E.B.	β	t	p
Block 1: Demographics Model Fit: $F(4, 4118) = 139.11, p < .001, R^2 = .12,$ F Change (4, 4118) = 139.11, $p < .001$					
IEP (0 = No, 1 = IEP)	-65.10	4.62	-0.21	-14.09	<0.001*
ELL (0 = No, 1 = ELL)	-27.72	3.34	-0.13	-8.29	<0.001*
FRL (0 = No, 1 = FRL)	-38.30	2.88	-0.21	-13.32	<0.001*
Gender (0 = M, 1 = F)	-0.78	2.61	-0.00	-0.30	0.766
Block 2: Demographics + Fall Score Model Fit: $F(5, 4117) = 304.92, p < .001, R^2 = .27,$ F Change (1, 4117) = 853.05, $p < .001$					
IEP (0 = No, 1 = IEP)	-41.95	4.28	-0.13	-9.80	<0.001*
ELL (0 = No, 1 = ELL)	-14.85	3.07	-0.07	-4.83	<0.001*
FRL (0 = No, 1 = FRL)	-17.45	2.71	-0.10	-6.43	<0.001*
Gender (0 = M, 1 = F)	1.38	2.38	0.01	0.58	0.561
Fall 2011 Test Score Scaled	0.37	0.01	0.42	29.21	<0.001*
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 4116) = 254.07, p < .001, R^2 = .27,$ F Change (1, 4116) = 0.15, $p = .698$					
IEP (0 = No, 1 = IEP)	-41.96	4.28	-0.13	-9.80	<0.001*
ELL (0 = No, 1 = ELL)	-14.84	3.08	-0.07	-4.83	<0.001*
FRL (0 = No, 1 = FRL)	-17.47	2.71	-0.10	-6.44	<0.001*
Gender (0 = M, 1 = F)	1.38	2.38	0.01	0.58	0.560
Fall 2011 Test Score Scaled	0.37	0.01	0.42	29.21	<0.001*
Phase (0 = P2, 1 = P1)	0.93	2.38	0.01	0.39	0.698

* $p < .05$.

Table 6. PASS-MC, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 4,123): All Regions

Area	Treatment (Phase 1)				Control (Phase 2)				F	p	g	PR
	n	M	SD	Adj. M	n	M	SD	Adj. M				
All	2,338	435.28	88.72	435.80	1,785	435.56	88.76	434.88	0.15	0.698	0.01	50
Not IEP	2,129	439.52	85.89	440.12	1,633	441.99	82.94	441.22	0.21	0.647	-0.01	49
IEP	209	392.08	104.38	390.08	152	366.49	116.11	369.23	3.86	0.050	0.19	58
Not ELL	1,801	444.95	82.60	445.10	1,367	445.28	80.80	445.09	0.00	0.997	0.00	50
ELL	537	402.87	100.18	405.64	418	403.77	104.85	400.21	0.80	0.370	0.05	52
Not FRL	922	466.01	67.46	466.07	725	459.22	73.53	459.14	5.14	0.024*	0.10	54
FRL	1,416	415.28	94.98	415.85	1,060	419.38	94.50	418.62	0.66	0.416	-0.03	49
Male	1,181	435.63	93.02	436.30	898	433.04	93.11	432.17	1.37	0.242	0.04	52
Female	1,157	434.92	84.14	435.33	887	438.12	84.10	437.59	0.50	0.481	-0.03	49

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

* $p < .05$

Middle School Cohort PASS-MC Spring 2014 Results: All Regions

For the 2,168 students across the three regions in the middle school cohort, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS-MC scaled scores (Block 3) explained 40% of the total variance (R^2) in students' 2014 Spring scaled scores as shown in Table 7. The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of Spring 2014 PASS-MC achievement ($\beta = -0.02$, $t = -1.16$, $p = .246$).

Unlike the outcomes observed for the elementary cohort, the overall performance result for the ANCOVA analysis (i.e., the "All" group) shown in Table 8 was negative for middle school cohort Phase 1 students ($n = 1,036$, Adjusted Mean = 323.02) compared to middle school cohort Phase 2 students ($n = 1,132$, Adjusted Mean = 327.22), and was not statistically significant ($F(1, 2161) = 1.35$, $p = 0.246$, $g = -0.04$, $PR = 48$). Meanwhile, the effect size ($g = -0.04$) was not substantively important. Despite the substantively important advantage of Phase 1 students on the Fall 2011 baseline for the ELL subgroup, Phase 2 outperformed Phase 1 for all subgroups except Not FRL, where the difference favoring Phase 1 was neither statistically significant, nor substantively important. The effect size favoring the Phase 2 IEP subgroup ($g = -0.28$) was the only statistically significant and substantively important subgroup effect found, and indicated that the average Phase 1 student scored at the 39th percentile of the control group.

Table 7. PASS-MC, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students (N = 2,168): All Regions

Source	B	S.E.B.	β	t	p
Block 1: Demographics Model Fit: $F(4, 2163) = 119.79, p < .001, R^2 = .18,$ F Change (4, 2163) = 119.79, $p < .001$					
IEP (0 = No, 1 = IEP)	-78.52	7.12	-0.22	-11.03	<0.001*
ELL (0 = No, 1 = ELL)	-61.54	7.88	-0.16	-7.81	<0.001*
FRL (0 = No, 1 = FRL)	-55.00	4.38	-0.25	-12.55	<0.001*
Gender (0 = M, 1 = F)	4.18	4.26	0.02	0.98	0.326
Block 2: Demographics + Fall Score Model Fit: $F(5, 2162) = 290.82, p < .001, R^2 = .40,$ F Change (1, 2162) = 798.33, $p < .001$					
IEP (0 = No, 1 = IEP)	-41.14	6.23	-0.12	-6.60	<0.001*
ELL (0 = No, 1 = ELL)	-27.79	6.84	-0.07	-4.06	<0.001*
FRL (0 = No, 1 = FRL)	-19.95	3.95	-0.09	-5.05	<0.001*
Gender (0 = M, 1 = F)	7.89	3.65	0.04	2.17	0.030*
Fall 2011 Test Score Scaled	0.53	0.02	0.53	28.26	<0.001*
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 2161) = 242.62, p < .001, R^2 = .40,$ F Change (1, 2161) = 1.35, $p = .246$					
IEP (0 = No, 1 = IEP)	-40.98	6.23	-0.12	-6.58	<0.001*
ELL (0 = No, 1 = ELL)	-27.97	6.85	-0.07	-4.09	0.006**
FRL (0 = No, 1 = FRL)	-19.50	3.97	-0.09	-4.92	<0.001*
Gender (0 = M, 1 = F)	7.97	3.65	0.04	2.19	0.029*
Fall 2011 Test Score Scaled	0.53	0.02	0.53	28.28	<0.001*
Phase (0 = P2, 1 = P1)	-4.21	3.63	-0.02	-1.16	0.246

* $p < .05$.

Table 8. PASS-MC, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 2,168): All Regions

Area	Treatment (Phase 1)				Control (Phase 2)				F	p	g	PR
	n	M	SD	Adj. M	n	M	SD	Adj. M				
All	1,036	323.75	110.85	323.02	1,132	326.55	106.00	327.22	1.35	0.246	-0.04	48
Not IEP	925	336.20	102.35	335.39	1,018	335.06	100.27	335.80	0.01	0.912	-0.00	50
IEP	111	220.02	124.36	217.99	114	250.53	124.45	252.50	5.91	0.016*	-0.28	39
Not ELL	953	331.42	107.82	332.23	1,032	335.61	101.50	334.86	0.50	0.478	-0.03	49
ELL	83	235.75	107.78	224.46	100	232.97	106.64	242.34	1.39	0.240	-0.17	43
Not FRL	392	363.24	97.68	364.67	518	365.63	86.91	364.54	0.00	0.981	0.00	50
FRL	644	299.72	111.57	293.04	614	293.58	109.45	300.58	2.25	0.134	-0.07	47
Male	505	314.03	116.58	314.88	570	322.31	112.06	321.56	1.45	0.229	-0.06	48
Female	531	333.00	104.39	330.93	562	330.84	99.39	332.80	0.16	0.690	-0.02	49

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

* $p < 0.05$.

Houston Independent School District: Results for Spring 2014 PASS Multiple Choice

Houston Independent School District (HISD) Spring 2014 PASS Multiple Choice Key Findings for Phase 1

For all students combined (the “All” group) and the specified subgroups, the following outcomes favoring Phase 1 elementary cohort students were found on the Spring 2014 PASS multiple choice section. Students in the middle school cohort (currently 8th graders in 2013-2014) took the PASS-MC for the first time in Spring 2012, and are therefore not included in these analyses.

Not Economically Disadvantaged (FRL)

- Elementary Cohort: While the difference was not statistically significant, Phase 1 students scored higher than Phase 2 students, with the magnitude of the difference being nearly substantively important (i.e., educationally meaningful) ($g = 0.23$).

Fall 2011 to Spring 2014 PASS Results: HISD

PASS-Multiple Choice: HISD

Students in the middle school cohort (currently 8th graders in 2013-2014) took the PASS-MC for the first time in Spring 2012, and are therefore not included in these analyses.

Table 9 shows the final sample sizes employed in the elementary cohort analyses (currently 5th graders in 2013-2014) once students missing data on all 29 PASS-MC questions at either time point were excluded.

Table 9. PASS-MC, Spring 2014: Samples for the PASS-MC Analyses for the Elementary Cohort: HISD

Sample	Phase 1	Phase 2
Students available for the PASS-MC achievement analysis	691	506

To determine baseline equivalence on the Fall 2011 PASS-MC between Phase 1 and Phase 2 students included in the present analysis, 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. In addition, an effect size was also calculated as a measure of baseline equivalence.

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

With respect to the elementary cohort (Table 10), students in the aggregate (the “All” group) did not demonstrate a statistically significant difference by Phase in their baseline achievement levels ($t(1143.99) = 1.68, p = .09, g = 0.10, PR = 54$), and the effect size was not substantively important. When the outcomes for Female students were compared by Phase, there was a statistically significant difference in Fall 2011 PASS scores that favored Phase 1, but the effect size linked to the comparison did not meet WWC criteria for substantive importance (i.e., $g \geq 0.25$) ($t(602) = 2.05, p = .04, g = 0.17, PR = 57$). On the other hand, there was a statistically significant difference in Fall 2011 PASS scores for Not ELL students, and the effect size associated with the difference met the WWC threshold for substantive importance, favoring Phase 1 students ($t(577) = 3.17, p < .01, g = 0.27, PR = 61$). Therefore, the outcome for the Not ELL subgroup comparison should be interpreted in light of the substantively important difference in baseline achievement between Phase 1 and Phase 2 students.

Employing these Fall 2011 data as covariates to statistically adjust the outcomes for baseline differences in achievement, preliminary analyses were conducted on Spring 2014 PASS-MC scaled scores to determine differences between Phase 1 and Phase 2 students in the elementary cohort, with each student’s scaled score on the Spring 2014 PASS-MC used as the outcome measure. As these analyses were exploratory in nature, no corrections were made for multiple comparisons.

Table 10. Baseline Comparison of Fall 2011 PASS-MC Scaled Scores for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 2,193): HISD

Group	Treatment (Phase 1)			Control (Phase 2)			<i>t</i>	<i>g</i>	PR
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>			
Elementary Cohort									
All	691	299.00	103.70	506	289.34	93.65	1.68	0.10	54
Not IEP	662	302.00	102.59	482	292.29	91.74	1.68	0.10	54
IEP	29	230.52	107.56	24	230.21	112.80	0.10	0.00	50
Not ELL	357	327.27	105.78	222	299.08	101.18	3.17*	0.27	61
ELL	334	268.79	92.46	284	281.73	86.74	-1.78	-0.14	44
Not FRL	114	368.61	119.59	55	351.31	111.89	0.90	0.15	56
FRL	577	285.25	94.49	451	281.78	88.39	0.60	0.04	52
Male	356	294.96	107.61	237	291.92	93.68	0.36	0.03	51
Female	335	303.30	99.36	269	287.07	93.74	2.05*	0.17	57

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

* $p < .05$.

Elementary Cohort PASS-Multiple Choice Analyses: HISD

With respect to the cohort of 1,197 elementary students in Phase 1 ($n = 691$) and Phase 2 ($n = 506$) schools, a hierarchical or “block entry” multiple regression was conducted to determine whether groups of students within the cohort differed by Phase in their performance on the Spring 2014 PASS-MC assessment scaled score (see Table 11). In addition to this regression, a second set of analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (*g*) were also conducted on the outcomes for all students by Phase within the cohort, as well as for subgroups of these same students, categorized by their IEP status, ELL status, FRL status, and Gender (see Table 12).

Elementary Cohort PASS-Multiple Choice Spring 2014 Results: HISD

For the 1,197 students across the three regions in the elementary cohort, the hierarchical multiple regression that controlled for student’s demographic characteristics and their Fall 2011 PASS-MC scaled scores (Block 3) explained 23% of the total variance (R^2) in students’ 2014 Spring PASS-MC scores (see Table 11). The addition of the student’s Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of Spring 2014 PASS-MC achievement ($\beta = -0.03$, $t = -1.10$, $p = .274$).

While the overall (i.e., the “All” group) ANCOVA adjusted mean presented in Table 12 was higher for Phase 2 students ($n = 506$, Adjusted Mean = 425.05) compared to Phase 1 students ($n = 691$, Adjusted Mean = 419.56), it fell short of being statistically significant ($F(1, 1190) = 1.20$, $p = 0.274$, $g = -0.06$, $PR = 48$), and the effect size ($g = -0.06$) was not substantively important. Consistent with these overall outcomes, six subgroup ANCOVA analyses (Not IEP, Not ELL, ELL, FRL, Male and Female) were linked to negatively signed effects and favored Phase 2 students in the elementary cohort (see Table 12). Nevertheless, none of the subgroup comparisons indicated a statistically significant difference between the adjusted means or an effect size that was substantively important. While not statistically significant, the effect size for the Not FRL subgroup ($g = 0.23$), which favored Phase 1, nearly reached the substantively important threshold (i.e., ≥ 0.25).

Table 11. PASS-MC, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students (N = 1,197): HISD

Source	B	S.E.B.	β	t	p
Block 1: Demographics Model Fit: $F(4, 1192) = 17.74, p < .001, R^2 = .06,$ F Change (4, 1192) = 17.74, $p < .001$					
IEP (0 = No, 1 = IEP)	-79.31	13.27	-0.17	-5.98	<0.001*
ELL (0 = No, 1 = ELL)	-16.37	5.79	-0.09	-2.83	0.005*
FRL (0 = No, 1 = FRL)	-35.73	8.30	-0.13	-4.31	<0.001*
Gender (0 = M, 1 = F)	3.94	5.47	0.02	0.72	0.471
Block 2: Demographics + Fall Score Model Fit: $F(5, 1191) = 70.80, p < .001, R^2 = .23,$ F Change (1, 1191) = 267.16, $p < .001$					
IEP (0 = No, 1 = IEP)	-46.18	12.17	-0.10	-3.79	<0.001*
ELL (0 = No, 1 = ELL)	-4.50	5.29	-0.02	-0.85	0.395
FRL (0 = No, 1 = FRL)	-6.18	7.72	-0.02	-0.80	0.424
Gender (0 = M, 1 = F)	3.53	4.95	0.02	0.71	0.475
Fall 2011 Test Score Scaled	0.43	0.03	0.44	16.35	<0.001*
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 1190) = 59.21, p < .001, R^2 = .23,$ F Change (1, 1190) = 1.20, $p = .274$					
IEP (0 = No, 1 = IEP)	-46.42	12.17	-0.10	-3.81	<0.001*
ELL (0 = No, 1 = ELL)	-4.82	5.29	-0.03	-0.91	0.363
FRL (0 = No, 1 = FRL)	-6.60	7.73	-0.03	-0.85	0.393
Gender (0 = M, 1 = F)	3.27	4.95	0.02	0.66	0.510
Fall 2011 Test Score Scaled	0.43	0.03	0.44	16.36	<0.001*
Phase (0 = P2, 1 = P1)	-5.49	5.01	-0.03	-1.10	0.274

* $p < .05$.

Table 12. PASS-MC, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 1,197): HISD

Area	Treatment (Phase 1)				Control (Phase 2)				F	p	g	PR
	n	M	SD	Adj. M	n	M	SD	Adj. M				
All	691	421.67	98.39	419.56	506	422.16	94.69	425.05	1.20	0.274	-0.06	48
Not IEP	662	424.55	96.40	422.58	482	426.04	91.88	428.74	1.49	0.223	-0.07	47
IEP	29	356.10	120.49	390.08	24	344.25	116.89	369.23	0.20	0.661	0.17	57
Not ELL	357	434.97	90.43	430.17	222	431.93	87.51	439.65	2.07	0.150	-0.11	46
ELL	334	407.46	104.52	410.25	284	414.52	99.43	411.24	0.17	0.896	-0.01	50
Not FRL	114	465.42	75.00	463.37	55	439.18	104.80	443.44	2.66	0.105	0.23	59
FRL	577	413.03	100.20	412.10	451	420.08	93.30	421.27	2.81	0.094	-0.09	46
Male	356	416.80	103.71	415.49	237	422.67	98.43	424.63	1.46	0.227	-0.09	46
Female	335	426.85	92.28	423.91	269	421.71	91.45	425.37	0.05	0.828	-0.02	49

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

* $p < .05$

New Mexico Region: Results for Spring 2014 PASS Multiple Choice

New Mexico Spring 2014 PASS Multiple Choice Key Findings for Phase 1

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

ELL

- Elementary Cohort: Phase 1 students scored statistically significantly higher than Phase 2, and the difference was substantively important ($g = 0.45$).

Fall 2011 to Spring 2014 PASS Results: New Mexico

PASS-Multiple Choice: New Mexico

Table 13 shows the final sample sizes employed in the elementary cohort analyses (currently 5th graders in 2013-2014) once students missing data on all 29 PASS-MC questions at either time point were excluded.

Table 13. PASS-MC, Spring 2014: Samples for the PASS-MC Analyses for the Elementary Cohort: New Mexico

Sample	Phase 1	Phase 2
Students available for the PASS-MC achievement analysis	520	317

Table 14 shows the final sample sizes employed in the middle school cohort analyses (currently 8th graders in 2013-2014) once students missing all 29 PASS-MC questions at either time point were excluded.

Table 14. PASS-MC, Spring 2014: Samples for the PASS-MC Analyses for the Middle School Cohort: New Mexico

Sample	Phase 1	Phase 2
Students available for the PASS-MC achievement analysis	489	116

To determine baseline equivalence on the Fall 2011 PASS-MC between Phase 1 and Phase 2 students included in the present analysis, 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. In addition, an effect size was also calculated as a measure of baseline equivalence.

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

With respect to the elementary cohort (Table 15), students in the aggregate (the “All” group) did not demonstrate a statistically significant difference by Phase in their baseline achievement levels ($t(835) = 1.27, p = .21, g = 0.09, PR = 54$). At the same time, students not designated as Economically Disadvantaged (Not FRL) were the only subgroup that appeared to have a statistically significant difference in baseline achievement, with Phase 1 Not FRL students outperforming their Phase 2 counterparts, although based on the effect size (*g*), not to a substantively meaningful degree ($t(438) = 2.22, p = .03, g = 0.22, PR = 59$). Overall, there were no substantively important effect size differences for the elementary cohort, meaning there was baseline equivalence for all groups.

Table 15. Baseline Comparison of Fall 2011 PASS-MC Scaled Scores for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 837): New Mexico

Group	Treatment (Phase 1)			Control (Phase 2)			t	g	PR
	n	M	SD	n	M	SD			
Elementary Cohort									
All	520	327.65	106.47	317	317.81	112.51	1.27	0.09	54
Not IEP	455	334.07	105.93	273	324.02	112.63	1.21	0.09	54
IEP	65	282.71	99.93	44	279.30	104.92	0.17	0.03	51
Not ELL	449	339.29	102.86	280	325.59	111.61	1.69	0.13	55
ELL	71	254.08	99.91	37	258.97	102.67	-0.24	-0.05	48
Not FRL	272	364.94	97.80	168	342.22	113.83	2.22*	0.22	59
FRL	248	286.75	100.57	149	290.29	104.71	-0.33	-0.03	49
Male	262	333.02	112.40	158	319.66	114.44	1.17	0.12	55
Female	258	322.21	100.01	159	315.97	110.89	0.59	0.06	52

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

* $p < .05$.

With respect to students in the middle school cohort (Table 16), across-the-board differences in baseline achievement were observed when the scores of Phase 1 students were compared with those of Phase 2. All favoring Phase 1 students and proving to be both statistically significant and substantively important were comparisons involving 6th Grade cohort students in the aggregate ($t(603) = 10.9, p < .001, g = 1.12, PR = 87$), and subgroups of students by their Special Education status—Not IEP ($t(535) = 9.8, p < .001, g = 1.11, PR = 87$) and IEP ($t(66) = 3.6, p = .001, g = 0.92, PR = 82$)—their status as English Language Learners—Not ELL ($t(545) = 8.2, p < .001, g = 0.98, PR = 84$) and ELL ($t(56) = 3.5, p = .001, g = 0.91, PR = 82$)—being identified as Economically Disadvantaged—FRL ($t(410) = 8.6, p < .001, g = 0.96, PR = 83$)—and their gender—Male ($t(285) = 7.0, p < .001, g = 1.02, PR = 85$) and Female ($t(316) = 8.3, p < .001, g = 1.22, PR = 89$). For the subgroup of students not identified as Economically Disadvantaged – (Not FRL) ($t(5.1) = 1.3, p = .26, g = 0.82, PR = 79$), the difference was not statistically significant, but was substantively important ($g = 0.82$).

Employing these Fall 2011 data as covariates to statistically adjust the outcomes for baseline differences in achievement, preliminary analyses were conducted on Spring 2014 PASS-MC scaled scores to determine differences between Phase 1 and Phase 2 students in the elementary and middle school cohorts, with each student's scaled score on the Spring 2014 PASS-MC used as the outcome measure. As these analyses were exploratory in nature, no corrections were made for multiple comparisons. Additionally, some violations in the ANCOVA assumptions were observed for subgroup comparisons.

Table 16. Baseline Comparison of Fall 2011 PASS-MC Scaled Scores for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 605): New Mexico

Group	Treatment (Phase 1)			Control (Phase 2)			t	g	PR
	n	M	SD	n	M	SD			
Middle School Cohort									
All	489	372.00	100.03	116	258.13	106.84	10.88*	1.12	87
Not IEP	443	381.95	96.55	94	273.64	101.13	9.80*	1.11	87
IEP	46	276.15	81.11	22	191.86	107.39	3.60*	0.92	82
Not ELL	465	376.56	98.85	82	277.87	111.45	8.17*	0.98	84
ELL	24	283.71	81.29	34	210.53	77.35	3.48*	0.91	82
Not FRL	187	409.11	92.97	6	331.17	149.72	1.27	0.82	79
FRL	302	349.02	97.44	110	254.15	103.46	8.60*	0.96	83
Male	228	359.42	103.27	59	250.80	116.19	7.01*	1.02	85
Female	261	383.00	95.96	57	265.72	96.65	8.35*	1.22	89

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

* $p < .05$.

Elementary and Middle School Cohort PASS-Multiple Choice Analyses: New Mexico

With respect to the cohort of 837 elementary students in Phase 1 ($n = 520$) and Phase 2 ($n = 317$) schools and the cohort of 605 middle school students in Phase 1 ($n = 489$) and Phase 2 ($n = 116$) schools, hierarchical or “block entry” multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their performance on the Spring 2014 PASS-MC assessment scaled score (see Table 17 and Table 19). In addition to these regressions, a second set of analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (g) were also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their IEP status, ELL status, FRL status, and Gender (see Table 18 and Table 20).

Elementary Cohort PASS-Multiple Choice Spring 2014 Results: New Mexico

For the 837 New Mexico students in the elementary cohort, the hierarchical multiple regression that controlled for student’s demographic characteristics and their Fall 2011 PASS-MC scaled scores (Block 3) explained 35% of the total variance (R^2) in students’ 2014 Spring PASS-MC scores (see Table 17). The addition of the student’s Phase to the model added 1% to the percentage of variance explained, but Phase was not a statistically significant predictor of Spring 2014 PASS-MC achievement ($\beta = 0.05$, $t = 1.91$, $p = .057$).

While the overall (i.e., the “All” group) ANCOVA adjusted mean presented in Table 18 was higher for Phase 1 students ($n = 520$, Adjusted Mean = 430.59) compared to Phase 2 students ($n = 317$, Adjusted Mean = 419.66), it fell short of being statistically significant ($F(1, 830) = 3.65$, $p = 0.057$, $g = 0.11$, $PR = 54$), and the effect size ($g = 0.11$) was not substantively important. Consistent with these overall outcomes, all subgroup ANCOVA analyses (Not IEP, IEP, Not ELL, ELL, Not FRL, FRL, Male, and Female) were linked to positively signed effects and favored Phase 1 students in the elementary cohort (see Table 18). Nevertheless, only the following three comparisons indicated a statistically significant difference between the adjusted means. In addition to being statistically significant, the difference for the comparison of students designated as receiving ELL services (ELL), favoring the 71 Phase 1 students

over the 37 Phase 2 students ($F(1, 102) = 5.96, p = 0.016, g = 0.45, PR = 67$), was also substantively important ($g = 0.45$) and indicated that the Phase 1 students scored at the 67th percentile of the control group. The other two comparisons with statistically significant differences were found not to be substantively important. The comparison for students who were not designated as receiving Special Education Services (Not IEP), favoring the 455 Phase 1 students over the 273 Phase 2 students ($F(1, 722) = 3.96, p = 0.047, g = 0.12, PR = 55$), indicated that the average Phase 1 student scored at the 55th percentile of the control group. Meanwhile, the comparison for Male students, favoring the 262 Phase 1 students over the 158 Phase 2 students ($F(1, 414) = 4.02, p = 0.046, g = 0.17, PR = 57$), indicated that the average Phase 1 student scored at the 57th percentile of the control group. Overall, only one of the effect sizes for the ANCOVA analyses was large enough to be substantively important, and ranged from a low of 0.03 (IEP) to a high of 0.45 (ELL).

Table 17. PASS-MC, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students (N = 837): New Mexico

Source	B	S.E.B.	β	t	p
Block 1: Demographics Model Fit: $F(4, 832) = 47.90, p < .001, R^2 = .19,$ F Change (4, 832) = 47.90, $p < .001$					
IEP (0 = No, 1 = IEP)	-57.72	9.31	-0.20	-6.20	<0.001*
ELL (0 = No, 1 = ELL)	-74.05	9.48	-0.25	-7.81	<0.001*
FRL (0 = No, 1 = FRL)	-49.71	6.38	-0.25	-7.79	<0.001*
Gender (0 = M, 1 = F)	2.92	6.25	0.02	-0.47	0.640
Block 2: Demographics + Fall Score Model Fit: $F(5, 831) = 86.34, p < .001, R^2 = .34,$ F Change (1, 831) = 195.33, $p < .001$					
IEP (0 = No, 1 = IEP)	-34.15	8.55	-0.12	-4.00	<0.001*
ELL (0 = No, 1 = ELL)	-51.55	8.69	-0.18	-5.93	<0.001*
FRL (0 = No, 1 = FRL)	-25.67	6.00	-0.13	-4.28	<0.001*
Gender (0 = M, 1 = F)	8.33	5.64	0.04	1.48	0.140
Fall 2011 Test Score Scaled	0.39	0.03	0.43	13.98	<0.001*
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 830) = 72.79, p < .001, R^2 = .35,$ F Change (1, 830) = 3.65, $p = .057$					
IEP (0 = No, 1 = IEP)	-33.99	8.54	-0.12	-3.98	<0.001*
ELL (0 = No, 1 = ELL)	-52.16	8.68	-0.18	-6.01	<0.001*
FRL (0 = No, 1 = FRL)	-25.82	5.99	-0.13	-4.31	<0.001*
Gender (0 = M, 1 = F)	8.38	5.63	0.04	1.49	0.137
Fall 2011 Test Score Scaled	0.39	0.03	0.43	13.88	<0.001*
Phase (0 = P2, 1 = P1)	10.94	5.73	0.05	1.91	0.057

* $p < .05$.

Table 18. PASS-MC, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 837): New Mexico

Area	Treatment (Phase 1)				Control (Phase 2)				F	p	g	PR
	n	M	SD	Adj. M	n	M	SD	Adj. M				
All	520	431.74	98.26	430.59	317	417.78	99.09	419.66	3.65	0.057	0.11	54
Not IEP	455	438.47	95.07	437.55	273	424.19	95.42	425.72	3.96	0.047*	0.12	55
IEP	65	384.58	107.71	383.44	44	381.94	109.24	379.72	0.04	0.847	0.03	51
Not ELL	449	442.31	90.33	440.00	280	431.50	86.78	435.19	0.71	0.400	0.05	52
ELL	71	364.92	118.81	366.20	37	313.97	123.61	311.50	5.96	0.016*	0.45	67
Not FRL	272	461.19	76.18	457.91	168	443.11	84.15	448.41	2.06	0.152	0.12	55
FRL	248	399.44	109.17	400.58	149	389.22	106.88	387.33	1.89	0.170	0.12	55
Male	262	431.13	101.79	428.60	158	407.15	105.54	411.34	4.02	0.046*	0.17	57
Female	258	432.36	94.74	432.26	159	428.34	91.35	428.49	0.25	0.618	0.04	52

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

* $p < .05$

Middle School Cohort PASS-MC Spring 2014 Results: New Mexico

For the 605 New Mexico students in the middle school cohort, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS-MC scaled scores (Block 3) explained 41% of the total variance (R^2) in students' 2014 Spring scaled scores as shown in Table 19. The addition of the student's Phase to the model added 1% to the percentage of variance explained and Phase was a statistically significant predictor of Spring 2014 PASS-MC achievement ($\beta = -0.08$, $t = -2.18$, $p = .029$), with Phase 2 students scoring statistically significantly higher than Phase 1 students.

Unlike the outcomes observed for the elementary cohort, the overall performance result for the ANCOVA analysis (i.e., the "All" group) shown in Table 20 was negative for middle school cohort Phase 1 students ($n = 489$, Adjusted Mean = 316.25) compared to middle school cohort Phase 2 students ($n = 116$, Adjusted Mean = 339.46) and was statistically significant ($F(1, 598) = 4.77$, $p = 0.029$, $g = -0.20$, $PR = 42$). However the effect size ($g = -0.20$) was not substantively important. In addition, even with the substantively important advantage Phase 1 students had on the Fall 2011 baseline for students overall and each of the subgroups, Phase 2 outperformed Phase 1 for all subgroups. Four of the subgroups favoring the Phase 2 students were found to be both statistically significant and substantively important: the comparison between students who were designated as receiving IEP services (IEP) ($F(1, 62) = 4.59$, $p = 0.036$, $g = -0.57$, $PR = 28$) indicated that the average Phase 1 student scored at the 28th percentile of the control group. The comparison between students who were not designated as receiving ELL services (Not ELL) ($F(1, 541) = 6.26$, $p = 0.013$, $g = -0.26$, $PR = 40$) indicated that the average Phase 1 student scored at the 40th percentile of the control group, while the comparison between students who were not designated as Economically Disadvantaged (Not FRL) ($F(1, 187) = 5.39$, $p = 0.021$, $g = -0.78$, $PR = 22$) indicated that the average Phase 1 student scored at the 22nd percentile of the control group. Finally, the comparison between Male students ($F(1, 281) = 6.59$, $p = 0.011$, $g = -0.34$, $PR = 37$) indicated that the average Phase 1 student scored at the 37th percentile of the control group. It should be noted that for the Not FRL subgroup, only 6 students were available for the analysis in Phase 2. Small sample sizes were also an issue for Phase 2 students in the IEP subgroup, and for Phase 1 students in the ELL subgroup. Therefore, due to the small sample sizes for these three subgroups, and due to the substantively important advantage of Phase 1 students on the baseline (Fall 2011) test, the results for the middle school cohort should be treated with caution.

Table 19. PASS-MC, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students (N = 605): New Mexico

Source	B	S.E.B.	β	t	p
Block 1: Demographics Model Fit: $F(4, 600) = 37.79, p < .001, R^2 = .20,$ F Change (4, 600) = 37.79, $p < .001$					
IEP (0 = No, 1 = IEP)	-74.46	14.22	-0.20	-5.24	<0.001*
ELL (0 = No, 1 = ELL)	-76.60	15.31	-0.19	-5.00	<0.001*
FRL (0 = No, 1 = FRL)	-64.36	9.36	-0.26	-6.87	<0.001*
Gender (0 = M, 1 = F)	12.42	8.78	0.05	1.42	0.158
Block 2: Demographics + Fall Score Model Fit: $F(5, 599) = 80.53, p < .001, R^2 = .40,$ F Change (1, 599) = 201.11, $p < .001$					
IEP (0 = No, 1 = IEP)	-28.48	12.74	-0.08	-2.24	0.026*
ELL (0 = No, 1 = ELL)	-36.39	13.56	-0.09	-2.68	0.007*
FRL (0 = No, 1 = FRL)	-26.61	8.53	-0.11	-3.12	0.002*
Gender (0 = M, 1 = F)	6.81	7.61	0.03	0.89	0.372
Fall 2011 Test Score Scaled	0.55	0.04	0.52	14.18	<0.001*
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 598) = 68.33, p < .001, R^2 = .41,$ F Change (1, 598) = 4.77, $p = .029$					
IEP (0 = No, 1 = IEP)	-27.21	12.71	-0.07	-2.14	0.033*
ELL (0 = No, 1 = ELL)	-43.06	13.86	-0.11	-3.11	0.002**
FRL (0 = No, 1 = FRL)	-29.22	8.59	-0.12	-3.40	0.001*
Gender (0 = M, 1 = F)	6.53	7.59	0.03	0.86	0.390
Fall 2011 Test Score Scaled	0.57	0.04	0.54	14.25	<0.001*
Phase (0 = P2, 1 = P1)	-23.20	10.62	-0.08	-2.18	0.029*

* $p < .05$

Table 20. PASS-MC, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 605): New Mexico

Area	Treatment (Phase 1)				Control (Phase 2)				F	p	g	PR
	n	M	SD	Adj. M	n	M	SD	Adj. M				
All	489	333.14	112.32	316.25	116	268.28	122.07	339.46	4.77	0.029*	-0.20	42
Not IEP	443	344.24	103.81	330.17	94	277.28	122.43	343.59	1.47	0.225	-0.12	45
IEP	46	226.26	134.39	203.42	22	229.82	115.37	277.58	4.59	0.036*	-0.57	28
Not ELL	465	337.88	110.01	327.94	82	300.10	111.87	356.49	6.26	0.013*	-0.26	40
ELL	24	241.21	119.29	205.75	34	191.53	112.38	216.56	0.11	0.736	-0.09	46
Not FRL	187	372.91	95.80	371.44	6	400.17	73.77	446.01	5.39	0.021*	-0.78	22
FRL	302	308.51	114.85	290.85	110	261.08	120.26	309.58	2.44	0.119	-0.16	44
Male	228	316.68	120.23	298.30	59	269.53	137.86	340.52	6.59	0.011*	-0.34	37
Female	261	347.52	103.02	332.82	57	266.98	104.48	334.29	0.01	0.914	-0.01	49

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

* $p < 0.05$.

North Carolina Region: Results for Spring 2014 PASS Multiple Choice

North Carolina Spring 2014 PASS Multiple Choice 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 multiple choice section.

IEP

- Elementary Cohort: Phase 1 had both a statistically significantly and substantively higher adjusted mean scaled score than Phase 2 in Spring 2014 ($g = 0.44$).

Fall 2011 to Spring 2014 PASS Results: North Carolina

PASS-Multiple Choice: North Carolina

Table 21 shows the final sample sizes employed in the elementary cohort analyses (currently 5th graders in 2013-2014) once students missing data on all 29 PASS-MC questions at either time point were excluded.

Table 21. PASS-MC, Spring 2014: Samples for the PASS-MC Analyses for the Elementary Cohort: North Carolina

Sample	Phase 1	Phase 2
Students available for the PASS-MC achievement analysis	1,127	962

Table 22 shows the final sample sizes employed in the middle school cohort analyses (currently 8th graders in 2013-2014) once students missing all 29 PASS-MC questions at either time point were excluded.

Table 22. PASS-MC, Spring 2014: Samples for the PASS-MC Analyses for the Middle School Cohort: North Carolina

Sample	Phase 1	Phase 2
Students available for the PASS-MC achievement analysis	547	1,016

To determine baseline equivalence on the Fall 2011 PASS-MC between Phase 1 and Phase 2 students included the present analysis, 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. In addition, an effect size was also calculated as a measure of baseline equivalence.

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

With respect to the elementary cohort (Table 23), students in the aggregate (the “All” group) demonstrated a statistically significant difference by Phase in their baseline achievement levels ($t(2087) = -3.28, p = 0.001, g = -0.14, PR = 44$), favoring Phase 2, 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, those who were not classified as receiving Special Education services ($t(1888) = -3.43, p = 0.001, g = -0.16, PR = 44$), those who were not English Language Learners ($t(1958) = -2.81, p = 0.005, g = -0.13, PR = 45$), those who were Economically Disadvantaged ($t(1049) = -2.33, p = .020, g = -0.15, PR = 44$) and those who were female

($t(1021) = -2.67, p = 0.008, g = -0.17, PR = 43$). Overall, there were no substantively important effect size differences for the elementary cohort, meaning there was baseline equivalence for all groups.

Table 23. Baseline Comparison of Fall 2011 PASS-MC Scaled Scores for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 2,089): North Carolina

Group	Treatment (Phase 1)			Control (Phase 2)			<i>t</i>	<i>g</i>	PR
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>			
Elementary Cohort									
All	1,127	312.8	96.32	962	326.4	92.84	-3.28**	-0.14	44
Not IEP	1,012	318.0	96.40	878	332.8	90.61	-3.43**	-0.16	44
IEP	115	267.4	83.14	84	259.9	90.19	0.61	0.09	53
Not ELL	995	320.3	95.40	865	332.6	93.33	-2.81**	-0.13	45
ELL	132	256.5	84.11	97	271.4	66.85	-1.44	-0.19	42
Not FRL	536	345.1	89.60	502	353.8	89.92	-1.56	-0.10	46
FRL	591	283.5	92.85	460	296.6	86.64	-2.33*	-0.15	44
Male	563	316.1	97.74	503	327.6	94.80	-1.95	-0.12	45
Female	564	309.5	94.86	459	325.2	90.73	-2.67**	-0.17	43

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

* $p < .05$; ** $p < .01$.

Likewise, with respect to students in the middle school cohort (Table 24), there was a statistically significant difference in baseline achievement by Phase ($t(1561) = -2.28, p = 0.023, g = -0.15, PR = 44$) in the aggregate, also favoring Phase 2, but the effect size linked to this difference was not substantively important. At the same time, not ELL students were the only subgroup that appeared to have a statistically significant difference in baseline achievement, with Phase 2 non-ELL students outperforming their Phase 1 counterparts, although based on the effect size (*g*), not to a substantively meaningful degree ($t(1436) = -2.10, p = 0.036, g = -0.14, PR = 44$). On the other hand, although there was no statistically significant difference in Fall 2011 PASS scores for ELL students, the effect size associated with the difference met the WWC threshold for substantive importance, favoring Phase 1 students ($t(123) = 1.66, p = 0.099, g = 0.39, PR = 65$). Therefore, the outcome for the ELL subgroup comparison for the middle school cohort should be interpreted in light of the substantively important difference in baseline achievement between Phase 1 and Phase 2 students.

Employing these Fall 2011 data as covariates to statistically adjust the outcomes for baseline differences in achievement, preliminary analyses were conducted on Spring 2014 PASS-MC scaled scores to determine differences between Phase 1 and Phase 2 students in the elementary and middle school cohorts, with each student's scaled score on the Spring 2014 PASS-MC used as the outcome measure. As these analyses were exploratory in nature, no corrections were made for multiple comparisons. Additionally, some violations in the ANCOVA assumptions were observed for subgroup comparisons. Therefore, the subgroup results should be interpreted with the statistical issues in mind.

Table 24. Baseline Comparison of Fall 2011 PASS-MC Scaled Scores for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 1,563): North Carolina

Group	Treatment (Phase 1)			Control (Phase 2)			t	g	PR
	n	M	SD	n	M	SD			
Middle School Cohort									
All	547	357.8	104.60	1,016	370.6	107.20	-2.28*	-0.15	44
Not IEP	482	367.4	100.60	924	378.0	104.70	-1.83	-0.12	45
IEP	65	286.5	106.60	92	296.4	103.90	-0.58	-0.12	45
Not ELL	488	365.6	104.30	950	377.8	104.50	-2.10*	-0.14	44
ELL	59	293.5	82.91	66	267.4	91.59	1.66	0.39	65
Not FRL	205	403.7	97.55	512	409.2	91.70	-0.71	-0.07	47
FRL	342	330.3	98.99	504	331.5	107.70	-0.16	-0.01	49
Male	277	362.6	107.50	511	373.2	110.00	-1.30	-0.12	43
Female	270	352.9	101.50	505	368.1	104.30	-1.95	-0.18	43

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

* $p < .05$.

Elementary and Middle School Cohort PASS-Multiple Choice Analyses: North Carolina

With respect to the cohort of 2,089 elementary students in Phase 1 ($n = 1,127$) and Phase 2 ($n = 962$) schools and the cohort of 1,563 middle school students in Phase 1 ($n = 547$) and Phase 2 ($n = 1,016$) schools, hierarchical or “block entry” multiple regressions were conducted to determine whether groups of students within cohorts differed by Phase in their performance on the Spring 2014 PASS-MC assessment scaled score (see Table 25 and Table 27). In addition to these regressions, a second set of analyses intended to generate pairs of adjusted scaled score means and to compute the treatment effect sizes (g) were also conducted on the outcomes for all students by Phase within cohort, as well as for subgroups of these same students, categorized by their IEP status, ELL status, FRL status, and Gender (see Table 26 and Table 28).

Elementary Cohort PASS-Multiple Choice Spring 2014 Results: North Carolina

For the 2,089 students in the elementary cohort in the North Carolina 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 PASS-MC scores (see Table 25). The addition of the student’s Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of Spring 2014 PASS-MC achievement ($\beta = -0.02$, $t = -1.04$, $p = 0.298$).

While the overall (i.e., the “All” group) ANCOVA adjusted mean presented in Table 26 was higher for Phase 1 students ($n = 1,127$, Adjusted Mean = 448.1) compared to Phase 2 students ($n = 962$, Adjusted Mean = 445.1), it also fell short of being statistically significant ($F(1, 2082) = 1.09$, $p = 0.298$, $g = 0.06$, $PR = 52$), and the effect size ($g = 0.06$) was not substantively important. Consistent with these overall outcomes, six subgroup ANCOVA analyses (IEP, Not ELL, ELL, Not FRL, FRL, and Male) were linked to positively signed effects and favored Phase 1 students in the elementary cohort (see Table 26). Nevertheless, only one of the comparisons indicated both a statistically significant and substantively important difference between the adjusted means: specifically, the comparison favoring the 115 Phase 1 students over the 84 Phase 2 students who were identified as receiving special education services

($F(1, 193) = 5.75, p = 0.017, g = 0.44, PR = 67$), and indicated that the average Phase 1 student scored at the 67th percentile of the control group. Additionally, Phase 1 male students statistically significantly outperformed Phase 2 male students on the posttest, but the difference was not substantively important ($F(1, 1060) = 5.47, p = 0.020, g = 0.17, PR = 57$).

Table 25. PASS-MC, Spring 2014: Hierarchical Multiple Regression Summary for Elementary Cohort Students (N = 2,089): North Carolina

Source	B	S.E.B.	β	t	p
Block 1: Demographics Model Fit: $F(4, 2,084) = 82.48, p < 0.001, R^2 = 0.137,$ F Change (4,2084) = 82.48, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-63.00	5.40	-0.24	-11.67	< 0.001***
ELL (0 = No, 1 = ELL)	-25.55	5.33	-0.10	-4.80	< 0.001***
FRL (0 = No, 1 = FRL)	-35.22	3.33	-0.23	-10.57	< 0.001***
Gender (0 = M, 1 = F)	-4.31	3.17	-0.03	-1.36	0.173
Block 2: Demographics + Fall Score Model Fit: $F(5, 2,083) = 157.31, p < 0.001, R^2 = 0.274,$ F Change (1,2083) = 394.36, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-43.47	5.05	-0.16	-8.62	< 0.001***
ELL (0 = No, 1 = ELL)	-12.86	4.93	-0.05	-2.61	0.009**
FRL (0 = No, 1 = FRL)	-18.51	3.17	-0.12	-5.84	< 0.001***
Gender (0 = M, 1 = F)	-2.30	2.90	-0.01	-0.79	0.429
Fall 2011 Test Score Scaled	0.33	0.02	0.40	19.86	< 0.001***
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 2,082) = 131.28, p < 0.001, R^2 = 0.274,$ F Change (1,2082) = 1.086, $p = 0.298$					
IEP (0 = No, 1 = IEP)	-43.55	5.05	-0.17	-8.63	< 0.001***
ELL (0 = No, 1 = ELL)	-12.89	4.93	-0.05	-2.62	0.009**
FRL (0 = No, 1 = FRL)	-18.59	3.17	-0.12	-5.86	< 0.001***
Gender (0 = M, 1 = F)	-2.37	2.91	-0.02	-0.81	0.415
Fall 2011 Test Score Scaled	0.33	0.02	0.40	19.89	< 0.001***
Phase (0 = P2, 1 = P1)	-3.03	2.91	-0.02	-1.04	0.298

** $p < .01$; *** $p < .001$

Table 26. PASS-MC, Spring 2014: Subgroup Mean Comparison for Elementary Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 2,089): North Carolina

Area	n	Treatment (Phase 1)			N	Control (Phase 2)			F	p	g	PR
		M	SD	Adj. M		M	SD	Adj. M				
All	1,127	445.3	75.73	448.1	962	448.5	79.49	445.1	1.09	0.298	0.06	52
Not IEP	1,012	449.8	71.70	452.6	878	456.3	70.02	453.1	0.03	0.871	-0.01	50
IEP	115	405.4	96.23	403.3	84	366.8	118.15	369.6	5.75	0.017*	0.44	67
Not ELL	995	449.7	75.42	452.3	865	453.2	75.91	450.2	0.51	0.476	0.04	52
ELL	132	411.7	69.57	414.0	97	406.6	97.05	403.4	1.16	0.283	0.17	57
Not FRL	536	468.6	60.74	470.3	502	466.8	63.88	465.0	2.38	0.123	0.12	55
FRL	591	424.1	81.59	426.4	460	428.5	89.48	425.6	0.03	0.860	0.01	51
Male	563	449.6	78.30	452.6	503	446.1	83.80	442.8	5.47	0.020*	0.17	57
Female	554	440.9	72.88	443.7	459	451.1	74.49	447.7	1.02	0.312	-0.08	47

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

* $p < .05$

Middle School Cohort PASS-MC Spring 2014 Results: North Carolina

For the 1,563 students in the middle school cohort in the North Carolina region, the hierarchical multiple regression that controlled for student's demographic characteristics and their Fall 2011 PASS-MC scaled scores (Block 3) explained 41% of the total variance (R^2) in students' 2014 Spring scaled scores as shown in Table 27. The addition of the student's Phase to the model did not add to the percentage of variance explained, and Phase was not a statistically significant predictor of Spring 2014 PASS-MC achievement ($\beta = 0.03$, $t = 1.50$, $p = 0.135$).

Unlike the outcome observed for the elementary cohort, the overall performance result for the ANCOVA analysis (i.e., the "All" group) shown in Table 28 was negative for middle school cohort Phase 1 students ($n = 547$, Adjusted Mean = 322.8) compared to middle school cohort Phase 2 students ($n = 1,016$, Adjusted Mean = 329.2), and was also not statistically significant ($F(1, 1556) = 2.24$, $p = 0.135$, $g = -0.07$, $PR = 47$). The effect size ($g = -0.07$) was also not substantively important. Additionally, Phase 2 students outperformed Phase 1 students for all subgroups in the Middle School Cohort. The effect size favoring Phase 2 IEP students ($g = -0.52$) was the only subgroup outcome that was both statistically significant and substantively, and indicated that the average Phase 1 student scored at the 30th percentile of the control group. In addition, despite the substantively important advantage of Phase 1 students on the Fall 2011 baseline for the ELL subgroup, the outcomes favored Phase 2 ELL students on the Spring 2014 PASS, with an effect size ($g = -0.65$) that while not statistically significant, was substantively important, indicating that the average Phase 1 student scored at the 26th percentile of the control group.

Table 27. PASS-MC, Spring 2014: Hierarchical Multiple Regression Summary for Middle School Cohort Students (N = 1,563): North Carolina

Source	B	S.E.B.	β	t	p
Block 1: Demographics Model Fit: $F(4, 1,558) = 82.19, p < 0.001, R^2 = 0.174$ F Change (4, 1558) = 82.19, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-79.32	8.21	-0.23	-9.66	< 0.001***
ELL (0 = No, 1 = ELL)	-54.20	9.20	-0.14	-5.89	< 0.001***
FRL (0 = No, 1 = FRL)	-52.76	4.98	-0.25	-10.60	< 0.001***
Gender (0 = M, 1 = F)	0.91	4.86	0.00	0.19	0.852
Block 2: Demographics + Fall Score Model Fit: $F(5, 1,557) = 211.05, p < 0.001, R^2 = 0.404,$ F Change (1, 1557) = 600.06, $p < 0.001$					
IEP (0 = No, 1 = IEP)	-46.04	7.11	-0.13	-6.48	< 0.001***
ELL (0 = No, 1 = ELL)	-24.24	7.92	-0.06	-3.06	0.002**
FRL (0 = No, 1 = FRL)	-18.46	4.46	-0.09	-4.14	< 0.001***
Gender (0 = M, 1 = F)	8.15	4.14	0.04	1.97	0.049*
Fall 2011 Test Score Scaled	0.52	0.02	0.53	24.50	< 0.001
Block 3: Demographics + Fall Score + Phase Model Fit: $F(6, 1,556) = 176.39, p < 0.001, R^2 = 0.405,$ F Change (1, 1556) = 2.236, $p = 0.135$					
IEP (0 = No, 1 = IEP)	-45.79	7.11	-0.13	-6.44	< 0.001***
ELL (0 = No, 1 = ELL)	-23.70	7.92	-0.06	-2.99	0.003
FRL (0 = No, 1 = FRL)	-17.77	4.48	-0.08	-3.97	< 0.001***
Gender (0 = M, 1 = F)	8.17	4.14	0.04	1.97	0.049*
Fall 2011 Test Score Scaled	0.52	0.02	0.53	24.50	< 0.001***
Phase (0 = P2, 1 = P1)	6.48	4.33	0.03	1.50	0.135

* $p < .05$; ** $p < .01$; *** $p < .001$

Table 28. PASS-MC, Spring 2014: Subgroup Mean Comparison for Middle School Cohort Phase 1 (Treatment) and Phase 2 (Control) (N = 1,563): North Carolina

Area	Treatment (Phase 1)				Control (Phase 2)				F	p	g	PR
	n	M	SD	Adj. M	n	M	SD	Adj. M				
All	547	315.4	108.95	322.8	1,016	333.2	101.97	329.2	2.24	0.135	-0.07	47
Not IEP	482	328.8	100.53	335.0	924	340.9	95.88	337.7	0.40	0.529	-0.03	49
IEP	65	215.6	117.63	217.7	92	255.5	126.62	254.0	4.74	0.031*	-0.52	30
Not ELL	488	325.3	105.43	331.7	950	338.7	100.03	335.4	0.67	0.412	-0.04	48
ELL	59	233.5	103.74	228.7	66	254.3	97.71	258.6	2.95	0.088	-0.65	26
Not FRL	205	354.4	98.77	357.9	512	365.2	87.04	363.8	0.89	0.347	-0.09	46
FRL	342	292.0	108.18	293.2	504	300.7	105.76	299.9	1.27	0.259	-0.08	47
Male	277	311.9	113.66	318.7	511	328.4	107.19	324.7	0.85	0.357	-0.07	47
Female	270	319.0	103.98	327.0	505	338.1	96.28	333.7	1.39	0.239	-0.09	46

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

* $p < 0.05$.

References

- Ferguson, C. J. (2009). An effect size primer: a guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532-538. Retrieved from http://psychology.okstate.edu/faculty/jgrice/psyc3214/AnEffectSizePrimer_2009.pdf
- Hayes, A. F., & Cai, L. (2007). Using heteroscedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. *Behavior Research Methods*, 39(4), 709-722. Retrieved from <http://www.afhayes.com/public/BRM2007.pdf>
- 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