Online appendix for

Differentiated Accountability and Education Production:

Evidence from NCLB Waivers

Steven W. Hemelt

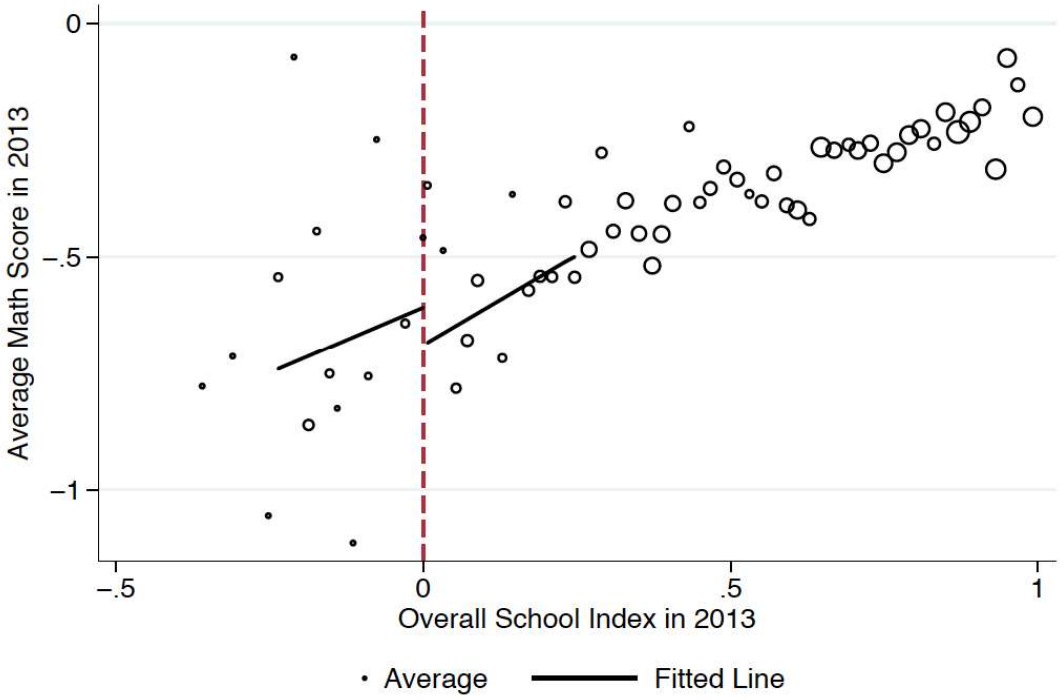
University of North Carolina at Chapel Hill

Brian A. Jacob

University of Michigan

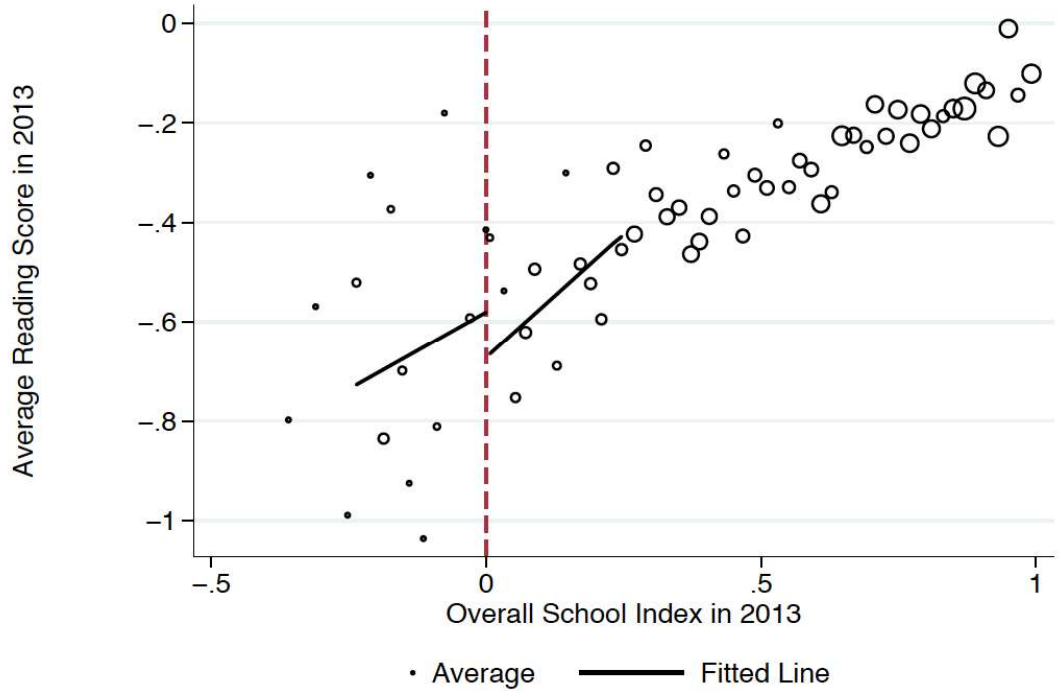
Appendix Figure A1. Baseline Equivalence: Priority Schools, 2013 Cohort

A. Average Math Scores



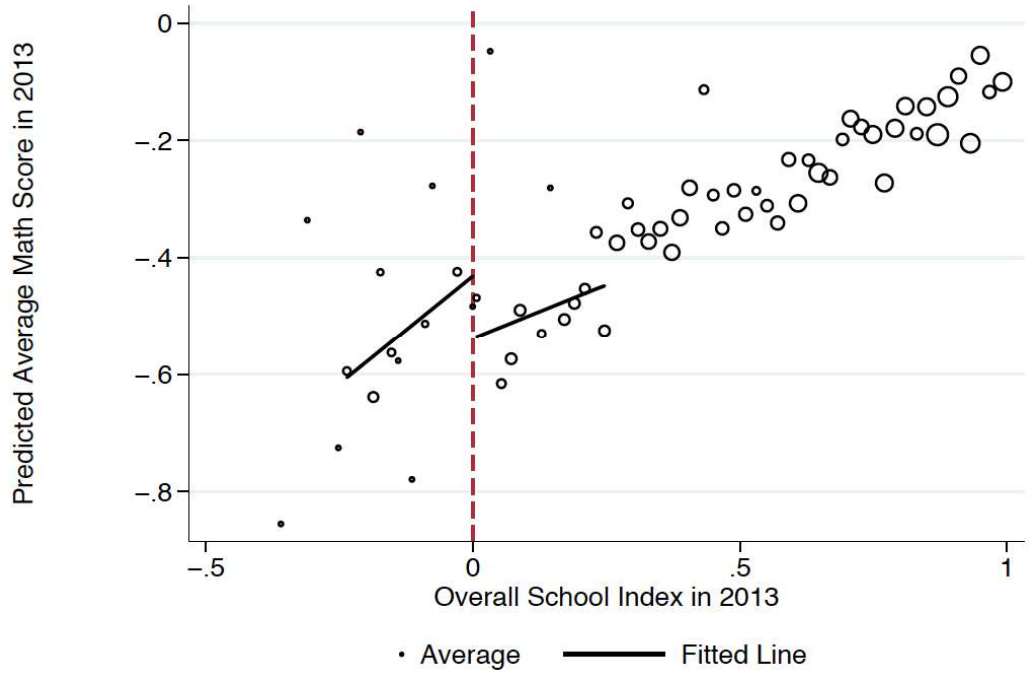
Linear: Estimate = 0.081 (0.133), N(schools) = 74, CM = −0.575, SD = 0.234
Nonparametric (mserd): Estimate = −0.099 [−0.733, 0.496], p = 0.706, bw = 0.135, N(schools) = 29, CM = −0.418

B. Average Reading Scores



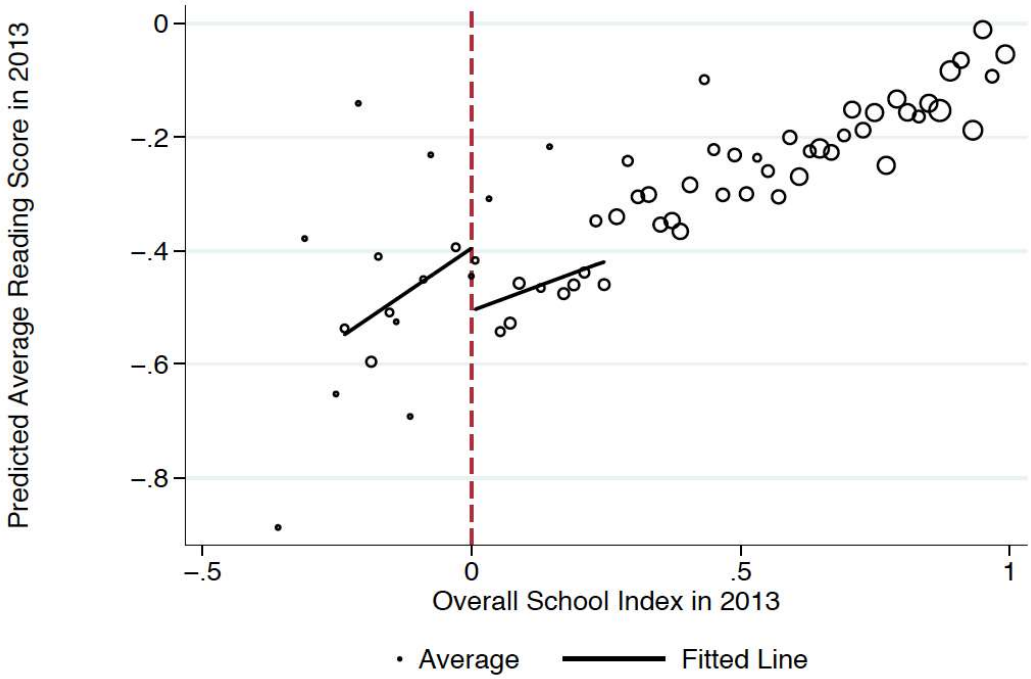
Linear: Estimate = 0.091 (0.115), N(schools) = 74, CM = −0.525, SD = 0.220
Nonparametric (mserd): Estimate = 0.053 [−0.404, 0.522], p = 0.803, bw = 0.157, N(schools) = 34, CM = −0.510

C. Predicted Average Math Scores



Linear: Estimate = 0.107 (0.110), N(schools) = 74, CM = −0.484, SD = 0.206
Nonparametric (mserd): Estimate = 0.063 [−0.301, 0.380], p = 0.821, bw = 0.187, N(schools) = 48, CM = −0.466

D. Predicted Average Reading Scores

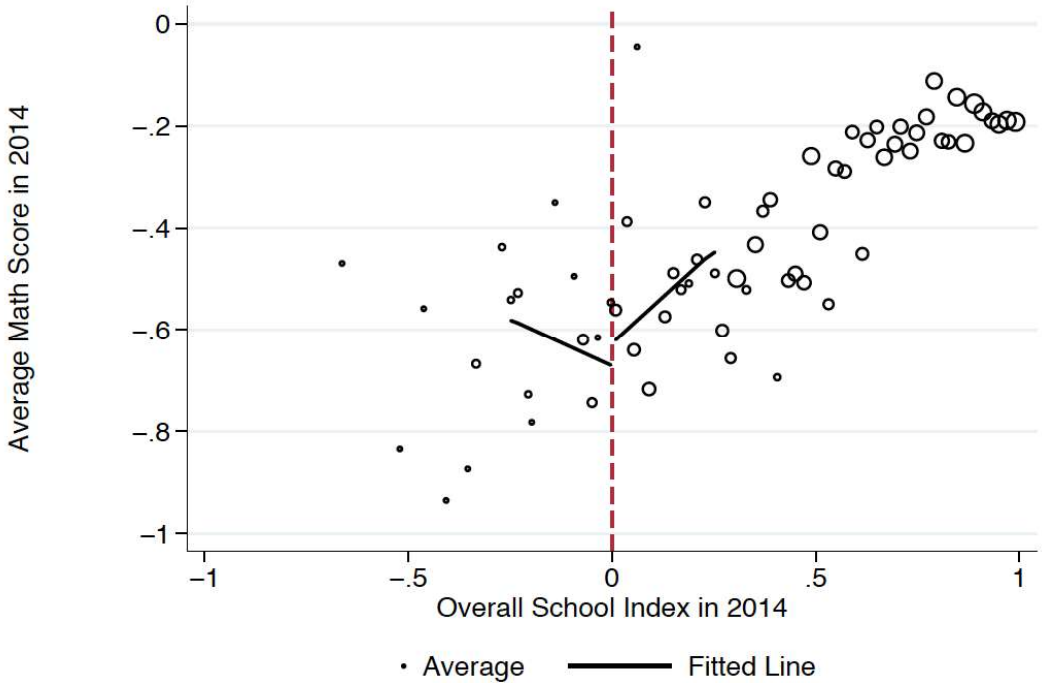


Linear: Estimate = 0.110 (0.097), N(schools) = 74, CM = −0.453, SD = 0.179
Nonparametric (mserd): Estimate = 0.067 [−0.280, 0.381], p = 0.763, bw = 0.170, N(schools) = 38, CM = −0.438

Notes: Analytic sample excludes schools identified as Priority in 2012. Parametric, linear specification is weighted by school enrollment. Nonparametric estimate is based on approach of Calonico et al. (2014a, 2014b, 2015). Predicted average scores come from regressions of test scores on the set of school characteristics reported in Table 1. CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

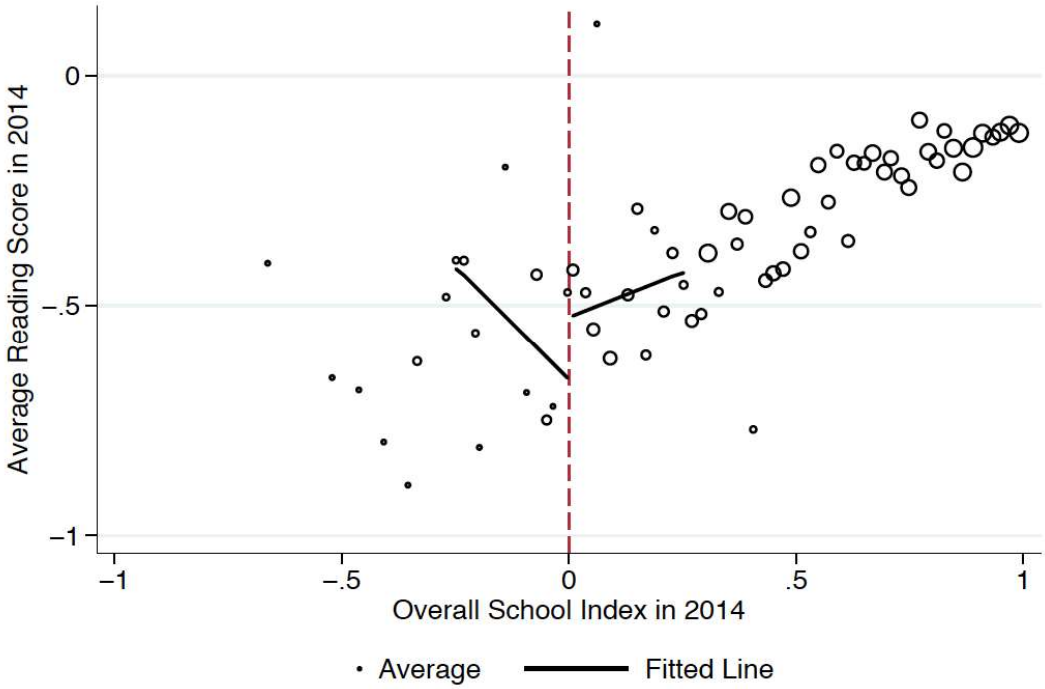
Appendix Figure A2. Baseline Equivalence: Priority Schools, 2014 Cohort

A. Average Math Score



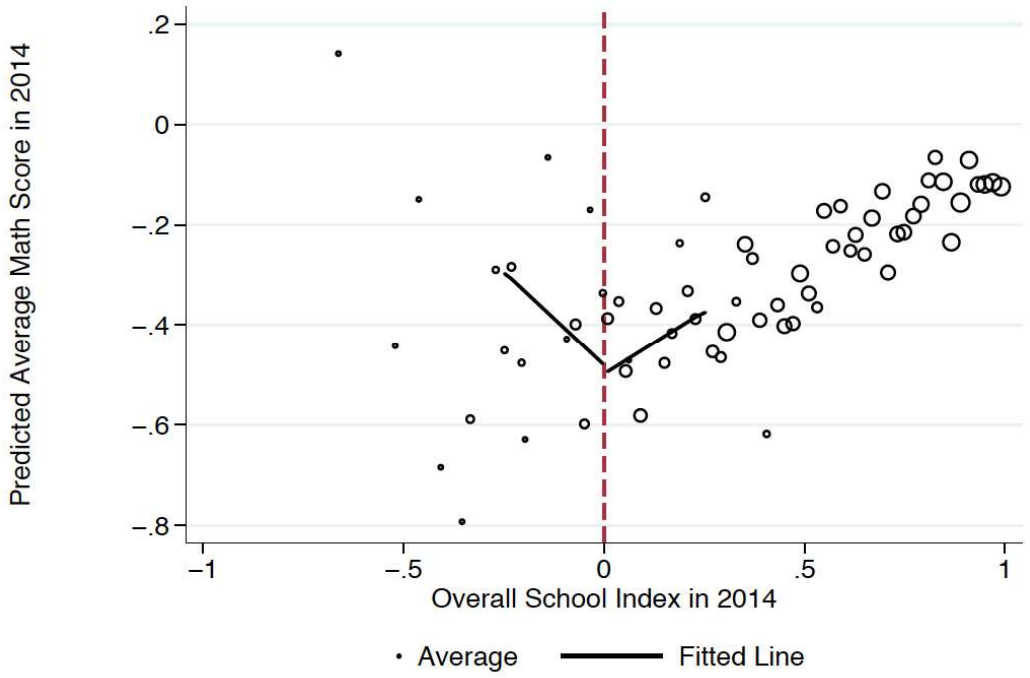
Linear: Estimate = −0.045 (0.094), N(schools) = 75, CM = −0.540, SD = 0.232
Nonparametric (mserd): Estimate = −0.111 [−0.391, 0.105], p = 0.259, bw = 0.179, N(schools) = 55, CM = −0.520

B. Average Reading Score



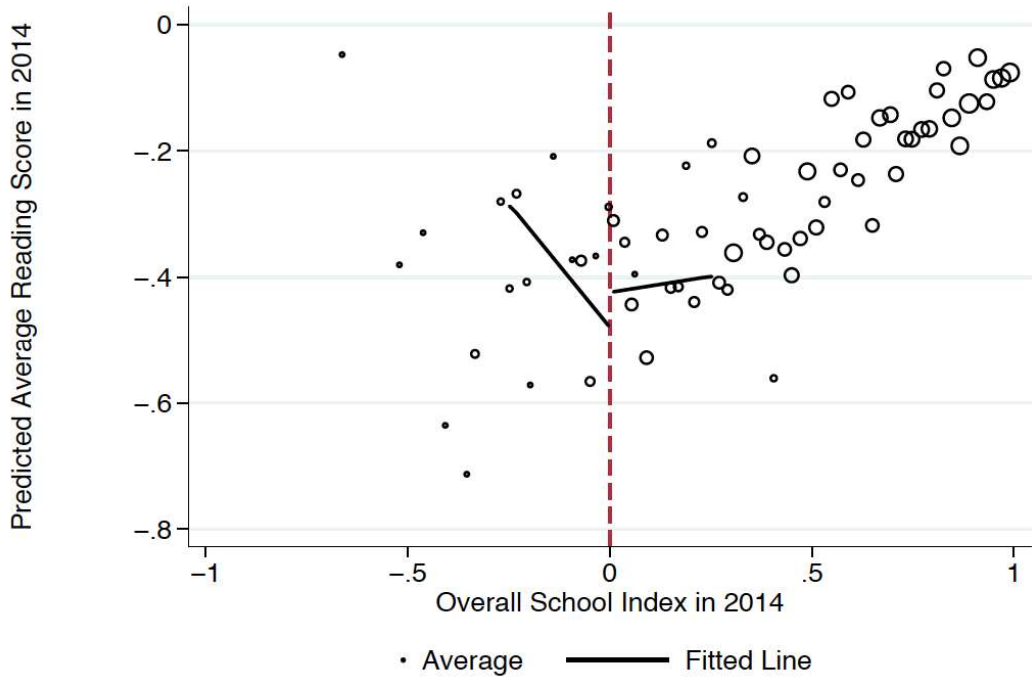
Linear: Estimate = −0.136 (0.111), N(schools) = 75, CM = −0.479, SD = 0.227
Nonparametric (mserd): Estimate = −0.175 [−0.534, 0.135], p = 0.243, bw = 0.191, N(schools) = 57, CM = −0.449

C. Predicted Average Math Score



Linear: Estimate = 0.018 (0.114), N(schools) = 75, CM = –0.439, SD = 0.222
Nonparametric (mserd): Estimate = –0.032 [–0.390, 0.256], p = 0.684, bw = 0.187, N(schools) = 56, CM = –0.394

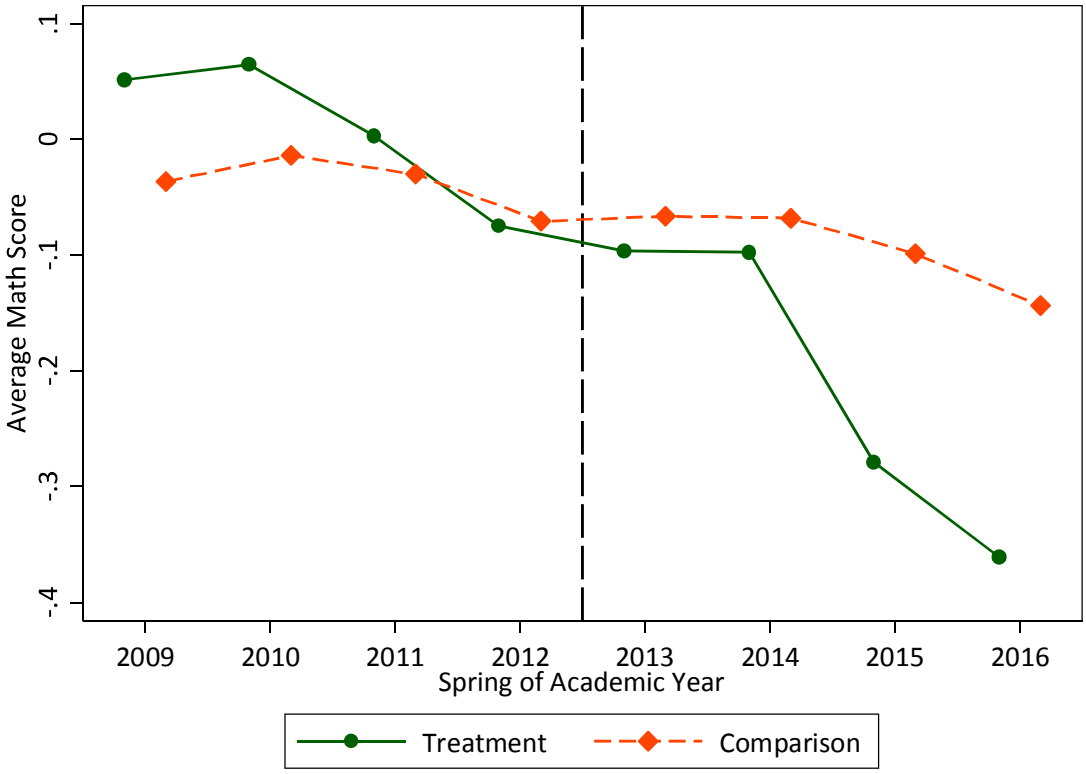
D. Predicted Average Reading Score



Linear: Estimate = –0.055 (0.097), N(schools) = 75, CM = –0.411, SD = 0.190
Nonparametric (mserd): Estimate = –0.069 [–0.397, 0.203], p = 0.527, bw = 0.178, N(schools) = 55, CM = –0.329

Notes: Analytic sample excludes schools identified as Priority in 2013 or 2012. Parametric, linear specification is weighted by school enrollment. Nonparametric estimate is based on approach of Calonico et al. (2014a, 2014b, 2015). Predicted average scores come from regressions of test scores on the set of school characteristics reported in Table 1. CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

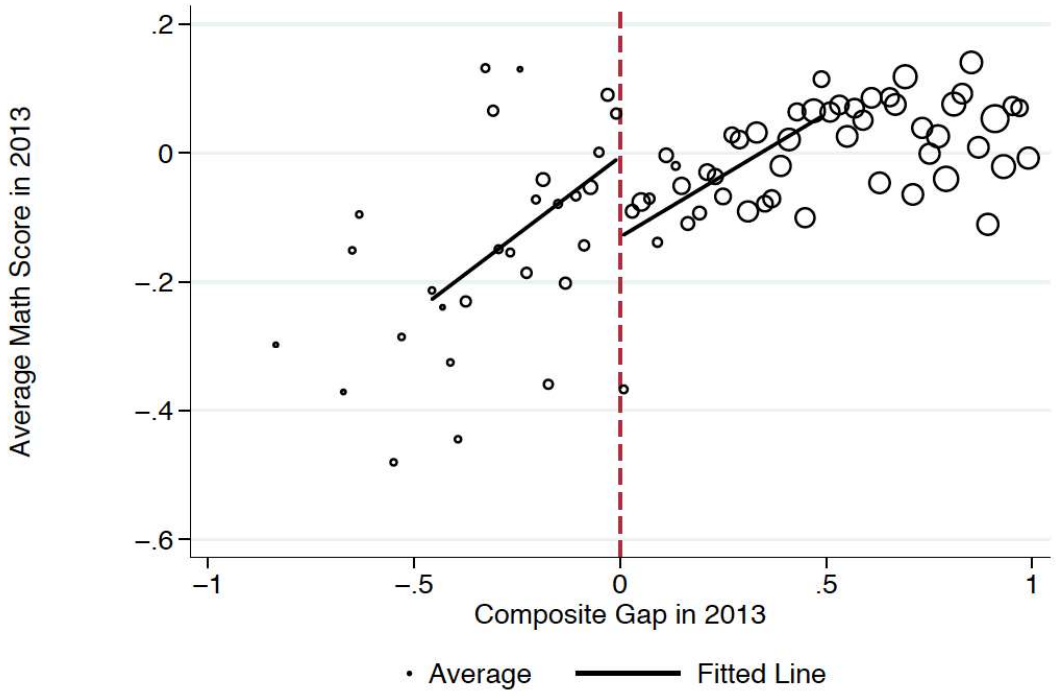
Appendix Figure A3. Trends in Average Math Performance, Priority Schools, 2012 Cohort



Notes: Points are coefficients from a regression of the outcome on year indicators. There are 30 Priority schools and 92 comparison schools represented in the plots. Comparison schools have running variable values that fall between 0.25 and 0.50, which correspond roughly to the 14th and 20th percentiles of the TTB rating distribution respectively.

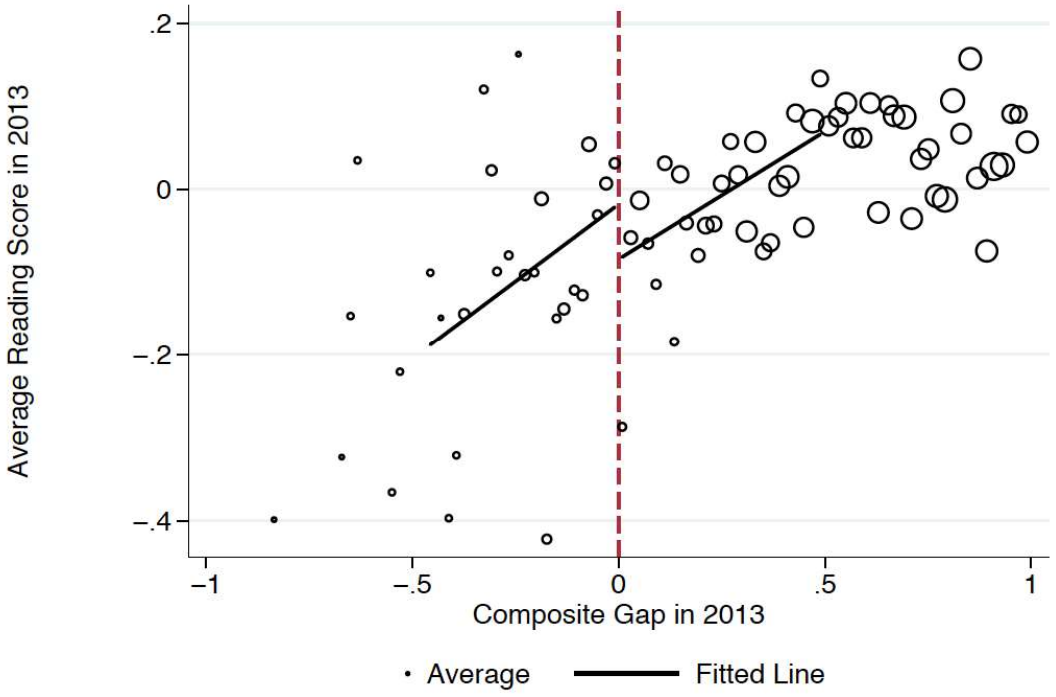
Appendix Figure A4. Baseline Equivalence: Focus Schools, 2013 Cohort

A. Average Math Scores



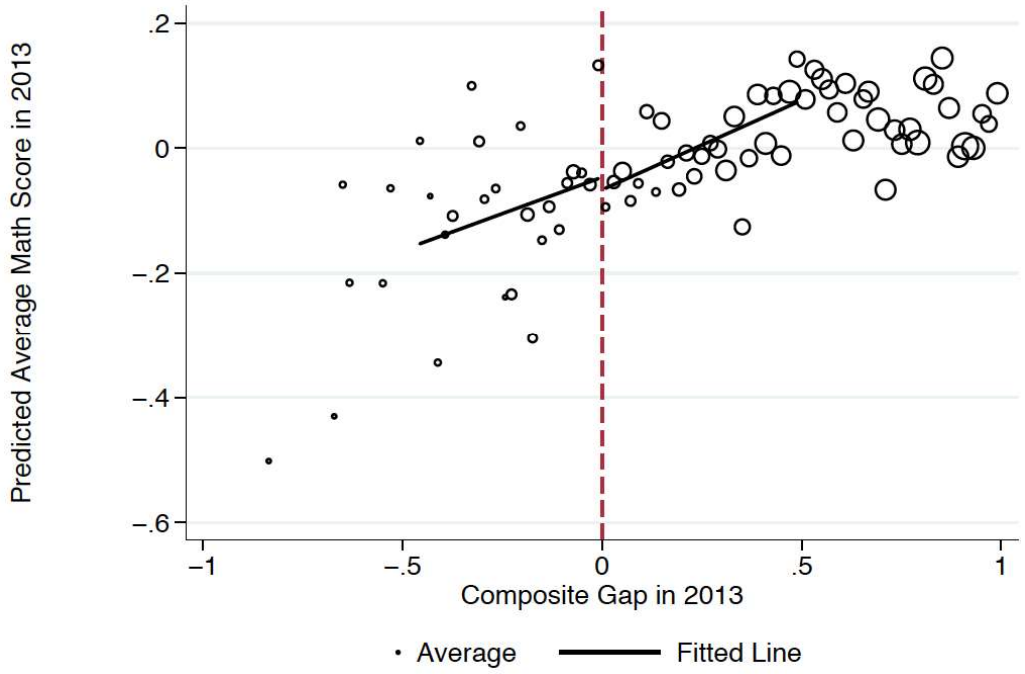
Linear: Estimate = 0.124** (0.057), N(schools) = 405, CM = −0.019, SD = 0.273
Nonparametric (mserd): Estimate = 0.260 [0.025, 0.571], p = 0.032, bw = 0.153, N(schools) = 99, CM = −0.186

B. Average Reading Scores



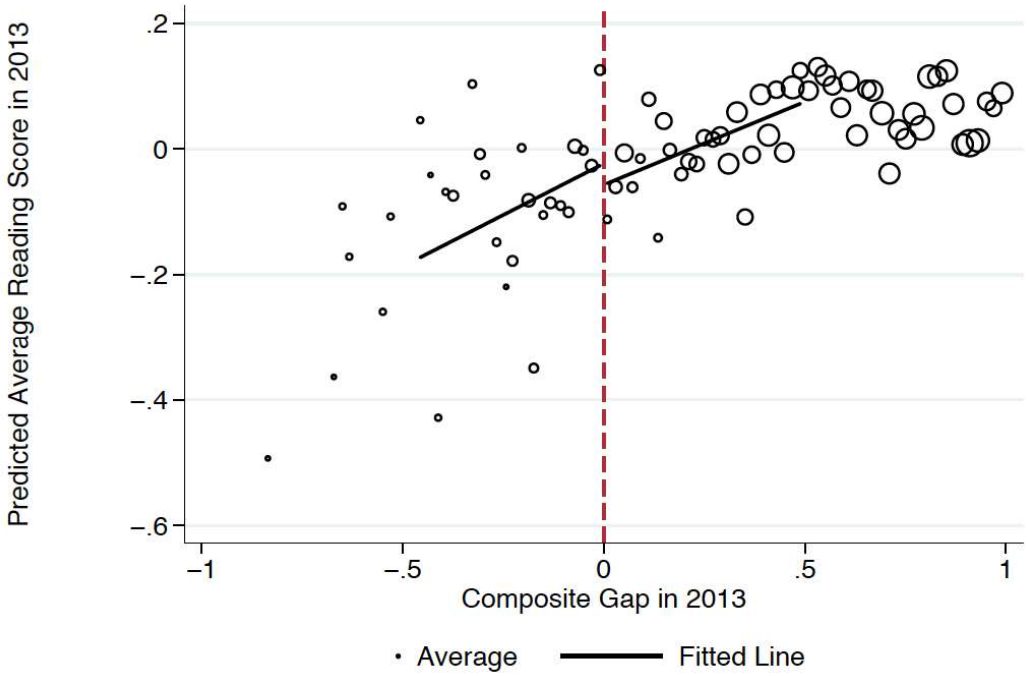
Linear: Estimate = 0.066 (0.052), N(schools) = 405, CM = 0.005, SD = 0.225
Nonparametric (mserd): Estimate = 0.135 [−0.050, 0.373], p = 0.135, bw = 0.172, N(schools) = 112, CM = −0.103

C. Predicted Average Math Scores



Linear: Estimate = 0.020 (0.059), N(schools) = 405, CM = 0.016, SD = 0.234
Nonparametric (mserd): Estimate = 0.124 [–0.075, 0.386], p = 0.185, bw = 0.205, N(schools) = 134, CM = –0.081

D. Predicted Average Reading Scores

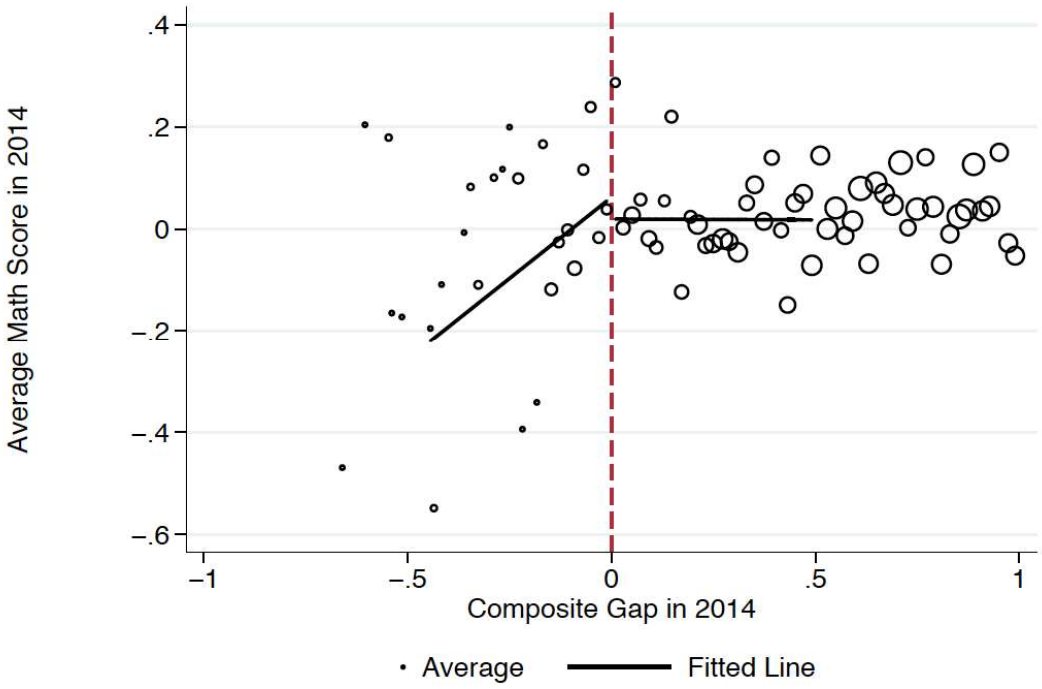


Linear: Estimate = 0.034 (0.056), N(schools) = 405, CM = 0.019, SD = 0.217
Nonparametric (mserd): Estimate = 0.136 [–0.047, 0.378], p = 0.126, bw = 0.200, N(schools) = 130, CM = –0.071

Notes: Analytic sample excludes schools identified as Focus or Priority in 2012. Parametric, linear specification is weighted by school enrollment. Nonparametric estimate is based on approach of Calonico et al. (2014a, 2014b, 2015). Predicted average scores come from regressions of test scores on the set of school characteristics reported in Table 1. CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

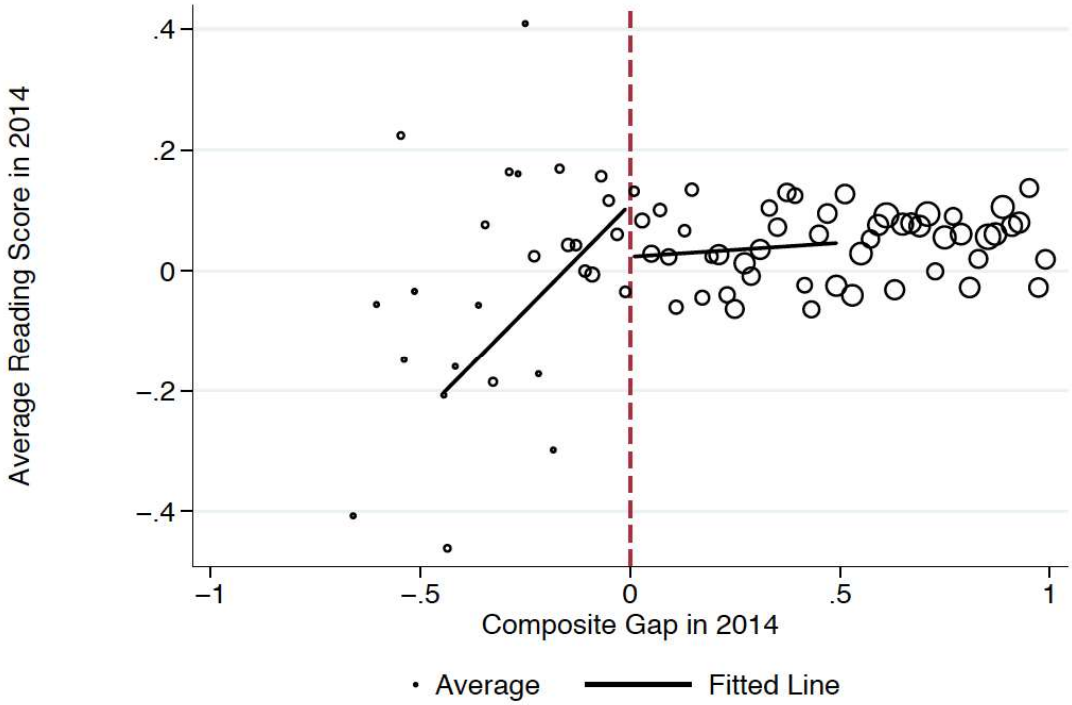
Appendix Figure A5. Baseline Equivalence: Focus Schools, 2014 Cohort

A. Average Math Scores



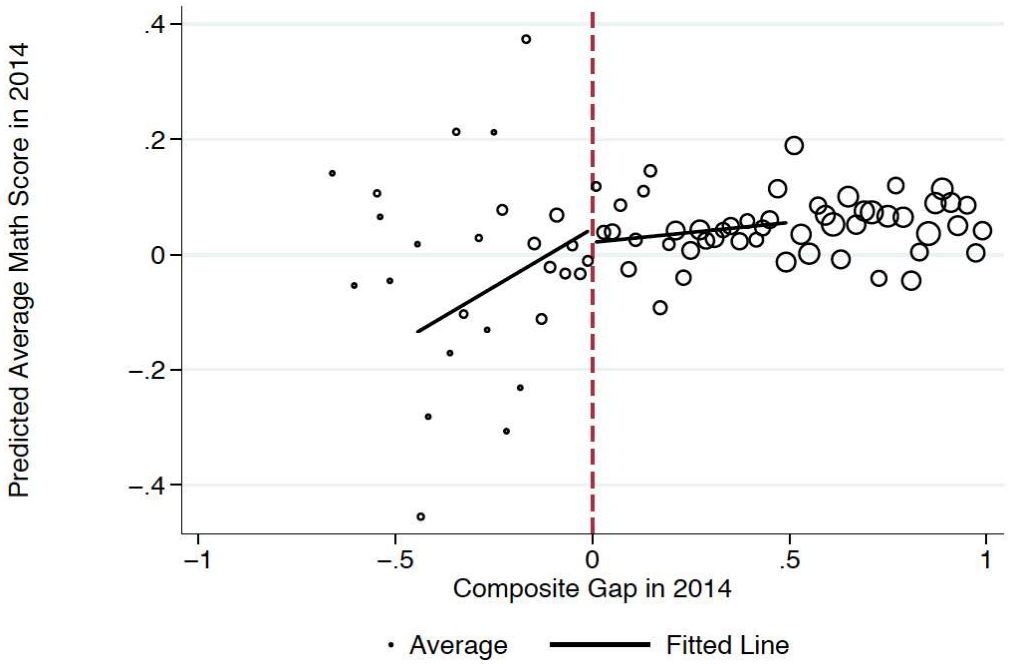
Linear: Estimate = 0.043 (0.079), N(schools) = 366, CM = 0.012, SD = 0.314
Nonparametric (mserd): Estimate = –0.047 [–0.406, 0.285], p = 0.733, bw = 0.177, N(schools) = 120, CM = 0.105

B. Average Reading Scores



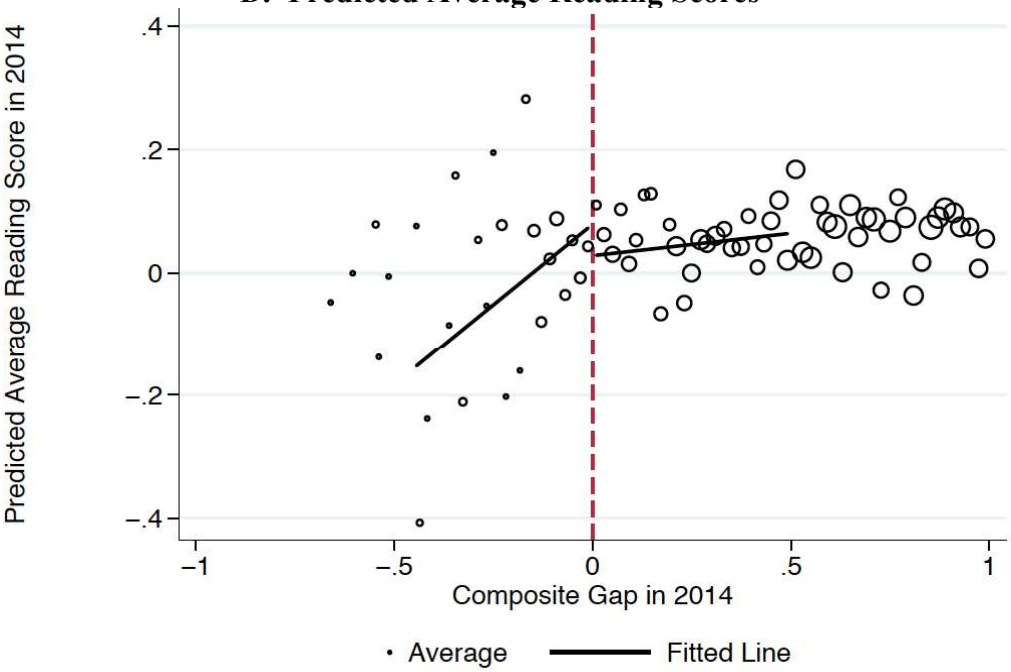
Linear: Estimate = 0.087 (0.056), N(schools) = 366, CM = 0.031, SD = 0.243
Nonparametric (mserd): Estimate = –0.080 [–0.318, 0.123], p = 0.388, bw = 0.187, N(schools) = 127, CM = 0.094

C. Predicted Average Math Scores



Linear: Estimate = 0.024 (0.059), N(schools) = 366, CM = 0.038, SD = 0.241
Nonparametric (mserd): Estimate = −0.097 [−0.295, 0.082], p = 0.268, bw = 0.187, N(schools) = 127, CM = 0.054

D. Predicted Average Reading Scores

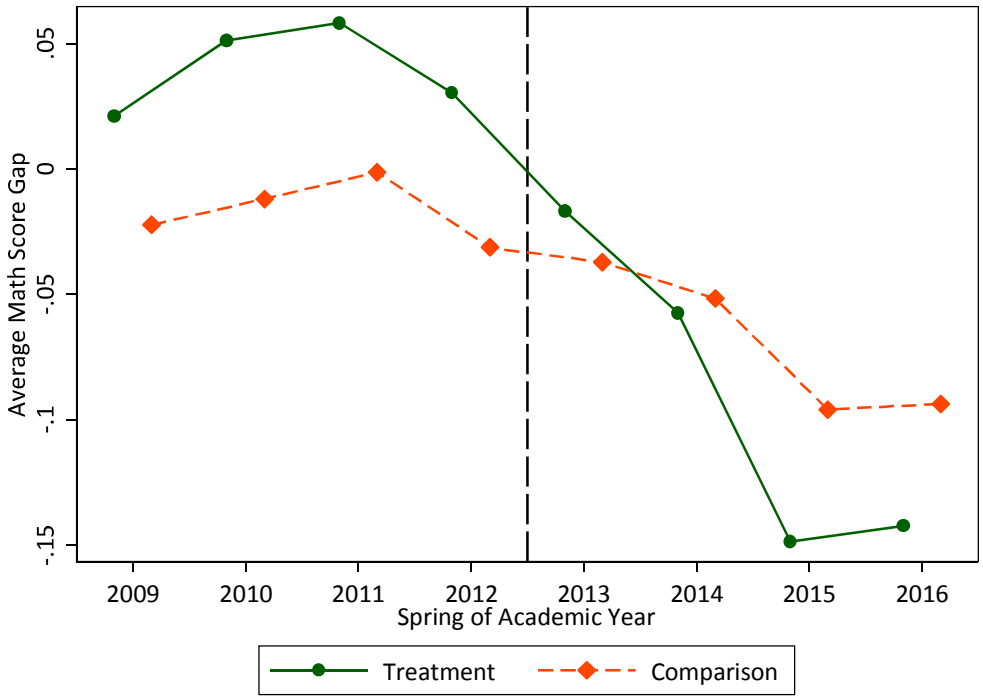


Linear: Estimate = 0.051 (0.052), N(schools) = 366, CM = 0.046, SD = 0.221
Nonparametric (mserd): Estimate = −0.061 [−0.235, 0.115], p = 0.500, bw = 0.187, N(schools) = 127, CM = 0.058

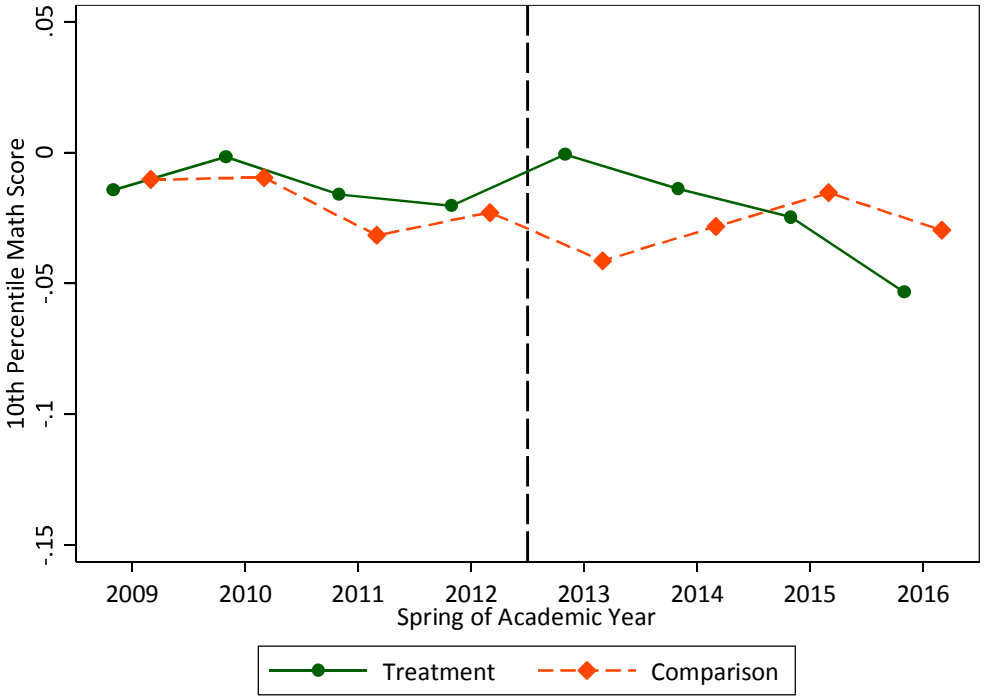
Notes: Analytic sample excludes schools identified as Focus or Priority in 2013 or 2012. Parametric, linear specification is weighted by school enrollment. Nonparametric estimate is based on approach of Calonico et al. (2014a, 2014b, 2015). Predicted average scores come from regressions of test scores on the set of school characteristics reported in Table 1. CM = control group mean; SD = standard deviation for control group; bw = bandwidth.

Appendix Figure A6. Trends in Math Performance, Focus Schools, 2012 Cohort

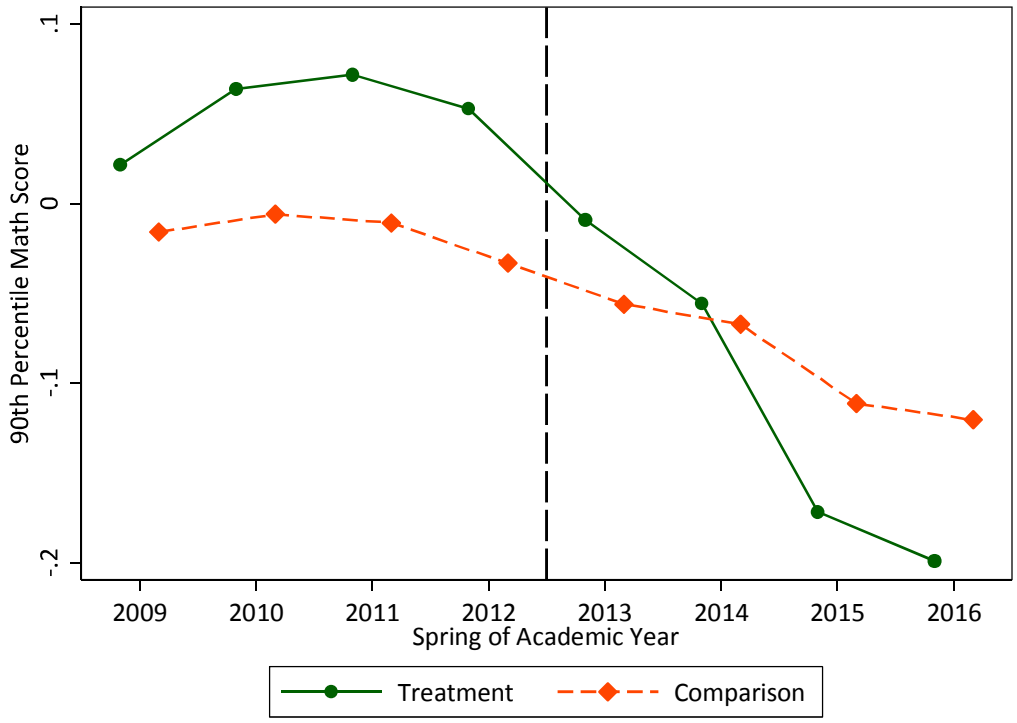
A. Within-School Math Gap Score



B. 10th Percentile Math Score



C. 90th Percentile Math Score



Notes: Points are coefficients from a regression of the outcome on year indicators. There are 160 Focus schools and 280 comparison schools represented in the plots. Comparison schools have running variable values that fall between 0.25 and 0.50, which correspond roughly to the 24th and 36th percentiles of the distribution of the achievement gap rating measure.

Appendix Table A1. Descriptive Statistics: Cohort of 2013

Variable	All K-8 Schools Mean (1)	Priority K-8 Schools Mean (2)	Focus K-8 Schools Mean (3)
Priority	0.030	1.000	--
Focus	0.127	--	1.000
Average Math Score (std)	0.040	-0.684	-0.099
	(0.428)	(0.316)	(0.230)
Average Reading Score (std)	0.042	-0.661	-0.083
	(0.357)	(0.297)	(0.221)
Priority Running Variable	1.495	-0.161	0.845
	(0.867)	(0.087)	(0.454)
Focus Running Variable	0.742	0.926	-0.227
	(0.710)	(0.920)	(0.182)
Student Characteristics			
Share black	0.149	0.623	0.151
Share white	0.732	0.260	0.722
Share Hispanic	0.071	0.095	0.079
Share Asian	0.034	0.014	0.031
Share economically disadvantaged	0.515	0.862	0.615
Share LEP	0.056	0.052	0.087
Share special education	0.147	0.193	0.153
Staff Characteristics			
Teacher-student ratio	0.082	0.086	0.088
Aide-student ratio	0.023	0.021	0.023
Share of teachers in first year at school	0.142	0.247	0.144
Average teacher experience (years)	14.9	14.5	14.6
	(2.5)	(4.0)	(2.6)
School Characteristics			
Total enrollment	464	394	432
	(192)	(131)	(185)
Elementary	0.680	0.769	0.598
Middle	0.245	0.038	0.289
Magnet	0.133	0.167	0.113
Urban	0.145	0.500	0.113
Suburban	0.458	0.458	0.423
Rural or Town	0.398	0.042	0.464
SIG Cohort III or IV	0.009	0.308	0.000
N(schools)	1805	26	97

Notes: Standard deviations of select continuous variables appear in parentheses below the means. Sample is limited to K-8, non-special-education, non-charter schools open as of the fall of 2013 with total enrollment in the baseline year of at least 50 students. Schools identified as persistently low achieving (PLA) prior to 2012 and schools that received SIG grants during the first two competitions (prior to 2012) are also excluded from the sample. The running variable for Priority schools is an index of prior level achievement, growth in achievement, and performance gaps; the running variable for Focus schools is a measure of the gap in performance between the top 30 percent and bottom 30 percent of students within a school. See text for additional details about running variables. LEP = limited English proficient; SIG = school improvement grant. Cohorts I and II of the SIG competition were identified before the first cohort of Priority and Focus schools. Priority and Focus schools were eligible to compete for SIG funding in Cohorts III and IV.

Appendix Table A2. Descriptive Statistics: Cohort of 2014

Variable	All K-8 Schools Mean (1)	Priority K-8 Schools Mean (2)	Focus K-8 Schools Mean (3)
Priority	0.039	1.000	--
Focus	0.128	--	1.000
Average Math Score (std)	0.041	-0.627	0.013
	(0.432)	(0.225)	(0.266)
Average Reading Score (std)	0.043	-0.560	0.034
	(0.363)	(0.265)	(0.201)
Priority Running Variable	1.467	-0.205	1.096
	(0.874)	(0.162)	(0.479)
Focus Running Variable	0.724	0.522	-0.173
	(0.729)	(1.052)	(0.147)
Student Characteristics			
Share black	0.152	0.484	0.119
Share white	0.727	0.357	0.789
Share Hispanic	0.073	0.113	0.054
Share Asian	0.034	0.036	0.027
Share economically disadvantaged	0.516	0.817	0.546
Share LEP	0.061	0.117	0.052
Share special education	0.146	0.193	0.150
Staff Characteristics			
Teacher-student ratio	0.083	0.090	0.081
Aide-student ratio	0.024	0.021	0.024
Share of teachers in first year at school	0.127	0.228	0.117
Average teacher experience (years)	15.0	15.6	14.4
	(2.6)	(3.0)	(2.3)
School Characteristics			
Total enrollment	460	465	466
	(192)	(184)	(202)
Elementary	0.676	0.781	0.635
Middle	0.242	0.094	0.311
Magnet	0.137	0.094	0.054
Urban	0.146	0.563	0.095
Suburban	0.456	0.375	0.365
Rural or Town	0.398	0.063	0.541
SIG Cohort III or IV	0.010	0.063	0.000
N(schools)	1789	32	74

Notes: Standard deviations of select continuous variables appear in parentheses below the means. Sample is limited to K-8, non-special-education, non-charter schools open as of the fall of 2014 with total enrollment in the baseline year of at least 50 students. Schools identified as persistently low achieving (PLA) prior to 2012 and schools that received SIG grants during the first two cohorts (prior to 2012) are also excluded from the sample. The running variable for Priority schools is an index of prior level achievement, growth in achievement, and performance gaps; the running variable for Focus schools is a measure of the gap in performance between the top 30 percent and bottom 30 percent of students within a school. See text for additional details about running variables. LEP = limited English proficient; SIG = school improvement grant. Cohorts I and II of the SIG competition were identified before the first cohort of Priority and Focus schools. Priority and Focus schools were eligible to compete for SIG funding in Cohorts III and IV.

Appendix Table A3. Effect of Treatment Status on School Closure

		Closed by 2012-2013			Closed by 2013-2014			Closed by 2014-2015			Closed by 2015-2016		
	N non-treated (treated) schools	Proportion of non-treated schools	Proportion of treated schools	ITT estimate (s.e.)	Proportion of non-treated schools	Proportion of treated schools	ITT estimate (s.e.)	Proportion of non-treated schools	Proportion of treated schools	ITT estimate (s.e.)	Proportion of non-treated schools	Proportion of treated schools	ITT estimate (s.e.)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Priority Analysis (Data window = +/- 0.25)													
2012 cohort	62(38)	0.06	0.21	0.011 (0.126)	0.16	0.29	0.007 (0.157)	0.18	0.34	0.146 (0.181)	0.18	0.39	0.191 (0.193)
2013 cohort	55(27)				0.07	0.15	0.167 (0.196)	0.09	0.18	0.218 (0.203)	0.13	0.19	0.239 (0.204)
2014 cohort	55(22)							0.02	0.05	0.055 (0.045)	0.09	0.09	0.049 (0.076)
Focus Analysis (Data window = +/- 0.50)													
2012 cohort	481(165)	0.01	0.01	-0.002 (0.008)	0.03	0.04	0.017 (0.020)	0.04	0.05	-0.003 (0.023)	0.06	0.07	0.009 (0.029)
2013 cohort	324(90)				0.02	0.01	-0.023 (0.015)	0.04	0.02	0.005 (0.035)	0.05	0.03	-0.018 (0.037)
2014 cohort	298(72)							0.01	0.00	-0.010 (0.015)	0.03	0.01	-0.025 (0.022)

Notes: Samples in column 1 are restricted to K-8, non-special-education, non-charter schools with total enrollment in the baseline year of at least 50 students. Subsequent cohorts in the Priority Analysis exclude schools previously identified as Priority schools. Subsequent cohorts in the Focus Analysis exclude schools previously identified as Priority or Focus schools. Coefficients in columns 4, 7, 10, and 13 represent effects from a simple linear, parametric RD specification estimated within the stated data window and weighted by school enrollment. Robust standard errors appear in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A4. Sensitivity of Achievement Estimates to Specification: Priority Schools

			1 year after		2 years after		3 years after		4 years after
Cohort			2012	2014	2013	2014	2012	2013	2012
Outcome and specification									
Data window	Polynomial	Covariates	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Outcome = Average Math Score (std)									
-1 to 1.5	cubic	yes	0.052	-0.174**	0.026	-0.085	0.055	-0.126	-0.131
			(0.070)	(0.080)	(0.099)	(0.086)	(0.093)	(0.126)	(0.158)
-0.25 to 0.25	linear	no	0.074	-0.237	0.038	-0.171	0.038	0.011	-0.050
			(0.143)	(0.164)	(0.178)	(0.178)	(0.204)	(0.171)	(0.226)
-0.25 to 0.25	linear	yes	0.003	-0.122	-0.109	-0.079	0.013	-0.160	-0.134
			(0.101)	(0.075)	(0.123)	(0.085)	(0.117)	(0.129)	(0.172)
Nonparametric	--	no	0.101	-0.295	-0.067	-0.236	0.180	-0.201	-0.094
(CCT)			(0.594)	(-1.419)	(-0.305)	(-0.979)	(0.866)	(-0.635)	(-0.382)
			[+/-0.23]	[+/-0.20]	[+/-0.16]	[+/-0.20]	[+/-0.24]	[+/-0.13]	[+/-0.24]
B. Outcome = Average Reading Score (std)									
-1 to 1.5	cubic	yes	0.010	-0.138	0.190	-0.079	0.146	-0.016	0.098
			(0.065)	(0.093)	(0.134)	(0.083)	(0.129)	(0.157)	(0.173)
-0.25 to 0.25	linear	no	-0.014	-0.328*	0.280	-0.267	0.176	0.081	0.152
			(0.151)	(0.183)	(0.217)	(0.187)	(0.248)	(0.188)	(0.263)
-0.25 to 0.25	linear	yes	-0.086	-0.173**	0.160	-0.151**	0.232	-0.035	0.066
			(0.100)	(0.077)	(0.109)	(0.073)	(0.145)	(0.135)	(0.147)
Nonparametric	--	no	-0.056	-0.338	-0.095	-0.282	0.209	-0.136	0.127
(CCT)			(-0.346)	(-1.357)	(-0.337)	(-0.993)	(0.889)	(-0.432)	(0.436)
			[+/-0.24]	[+/-0.20]	[+/-0.12]	[+/-0.18]	[+/-0.24]	[+/-0.12]	[+/-0.23]

Notes: The analytic sample for the 2012 cohort includes all students enrolled in K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of at least 50 students. For subsequent cohorts, we exclude schools identified as Priority in prior years since they leave the risk set for Priority treatment after initial designation for intervention. For a full list of covariates, please consult the text. All specifications are weighted by total school enrollment. Robust standard errors appear in parentheses below the parametric estimates. Z-statistics (bandwidths) appear in parentheses (brackets) below the nonparametric estimates *** p<0.01, ** p<0.05, * p<0.1. CCT = Calonico, Cattaneo, and Titiunik (2015).

Appendix Table A5. Sensitivity of Achievement Estimates to Specification: Focus Schools

			1 year after		2 years after		3 years after		4 years after
Cohort			2012	2014	2013	2014	2012	2013	2012
Outcome and specification									
Data window	Polynomial	Covariates	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Outcome = Math Gap Score									
-1 to 1.5	cubic	yes	-0.108***	-0.060	-0.006	-0.011	-0.047	0.013	-0.035
			(0.038)	(0.061)	(0.058)	(0.056)	(0.038)	(0.054)	(0.039)
-0.50 to 0.50	linear	no	-0.112***	-0.060	0.024	0.016	-0.028	0.022	-0.025
			(0.039)	(0.040)	(0.039)	(0.040)	(0.028)	(0.041)	(0.028)
-0.50 to 0.50	linear	yes	-0.083***	-0.051	0.012	0.030	-0.031	0.019	-0.033
			(0.029)	(0.040)	(0.035)	(0.041)	(0.028)	(0.039)	(0.028)
Nonparametric	--	no	-0.089**	-0.208***	0.039	-0.050	-0.010	0.082	0.000
(CCT)			(-2.192)	(-2.566)	(0.0445)	(-0.780)	(-0.319)	(1.140)	(0.007)
			[+/-0.48]	[+/-0.14]	[+/-0.15]	[+/-0.19]	[+/-0.49]	[+/-0.20]	[+/-0.49]
B. Outcome = Reading Gap Score									
-1 to 1.5	cubic	yes	-0.040	-0.057	0.003	-0.047	-0.007	0.057	0.003
			(0.032)	(0.054)	(0.056)	(0.051)	(0.037)	(0.040)	(0.034)
-0.50 to 0.50	linear	no	-0.025	-0.032	0.015	-0.012	0.001	0.027	-0.002
			(0.025)	(0.035)	(0.034)	(0.035)	(0.028)	(0.031)	(0.026)
-0.50 to 0.50	linear	yes	-0.022	-0.026	0.014	-0.005	-0.010	0.033	-0.015
			(0.025)	(0.036)	(0.034)	(0.038)	(0.026)	(0.032)	(0.025)
Nonparametric	--	no	-0.056*	-0.288***	-0.010	-0.072	0.021	0.037	0.03
(CCT)			(-1.950)	(-3.886)	(-0.153)	(-1.353)	(0.646)	(0.812)	(0.941)
			[+/-0.47]	[+/-0.11]	[+/-0.16]	[+/-0.23]	[+/-0.47]	[+/-0.25]	[+/-0.42]

Notes: The analytic sample for the 2012 cohort includes all students enrolled in K-8, non-special-education, non-charter schools open as of the fall of 2012 with total enrollment in the baseline year of at least 50 students. For subsequent cohorts, we exclude schools identified as Focus or Priority in prior years since they leave the risk set for Focus treatment after initial designation for intervention. For a full list of covariates, please consult the text. All specifications are weighted by total school enrollment. Robust standard errors appear in parentheses below the parametric estimates. Z-statistics (bandwidths) appear in parentheses (brackets) below the nonparametric estimates *** p<0.01, ** p<0.05, * p<0.1. CCT = Calonico, Cattaneo, and Titiunik (2015).

Appendix B. Calculation of Top-to-Bottom (TTB) Index and the Identification of Priority and Focus Schools in Michigan

Overview

The TTB is a performance index that was developed in order to implement a system of differentiated accountability and supports under Michigan's approved waiver from NCLB. The index ranks public schools on student performance in mathematics, reading, writing, science, and social studies, as well as graduation rates (for high schools). Performance in each tested subject (i.e., content area) is measured in three ways: level of achievement, growth in achievement over time, and the gap in achievement between the highest 30 percent and lowest 30 percent of students within each school.

Which schools get a TTB ranking?

Schools must have at least 30 full-academic-year (FAY) students in either the elementary/middle or high school span (or both) with test scores over two years in at least two subjects.[1] Traditional public schools as well as charter schools that meet this criterion are ranked.

Which students are included in a school's TTB ranking?

All FAY public school students with valid test scores on regular or alternate assessments within eligible schools are included. Thus, special education students who traditionally take alternate exams are included. Homeschooled and private school students with state test scores are not included.

How does Michigan calculate the overall TTB score?

We describe the rules that identified the first cohort of Priority and Focus schools in August of 2012. To calculate the TTB index for each school, the Michigan Department of Education (MDE) used students' test scores[2] over a minimum of two prior years.

After assigning each student to her main school for each academic year, the state calculated standardized scale scores for each student in each subject (and test type) in which she was tested by year and grade level. These standardized scores provide a measure of how well a student performed relative to her peers across the state who took the same test in the same year.

For each subject area in which a school had the requisite number of tested students over a period of two prior years, the state created a "content area index" that was a function of prior test scores, change in student performance, and the within-school gap between the top 30 percent and bottom 30 percent of students. Below, we describe how the state calculated each of these components for a given subject area (e.g., mathematics):

---

[1] Closed schools are provided a ranking if adequate historical data exist. A school classified as both an elementary/middle and high school has ranks calculated for both sets of grades and the final rank is an average of the two.

[2] Michigan's exams include the Michigan Educational Assessment Program (MEAP) tests given in mathematics, reading, writing, science, and social studies across grades 3 through 9; MEAP-Access, which is the alternate assessment given to students with special education needs; and the Michigan Merit Examination (MME), given in 11th grade, which includes the ACT.

- Two-year average scores (weight = 50%): For each school, the state calculated the average of student-level standardized scale scores in subject X in year t-1 (i.e., the prior year) and year t-2 (i.e., two years prior). Along with these means, the state counted the number of students tested in subject X in the school for each of those prior two years. Then the state calculated a weighted average of these two means and standardized the resulting average across comparable schools (i.e., elementary/middle or high schools). That value became the school-level standardized score for subject X.

- Growth in student performance (weight = 25%): The state used a two-pronged approach to assess improvements in performance. For a given school, the choice of approach for each subject/content area rested on whether students were tested in that subject over multiple years (e.g., math tests in grades 3, 4, and 5).

  For subjects in which adjacent-year testing occurred (e.g., math and reading in elementary and middle school grades), the state used a minimum of three years of historical data to classify students into different groups of "performance level changes," which were based on a student's movement across sub-categories of the four main proficiency-level categories: not proficient, partially proficient, proficient, and advanced.

  The state divided each proficiency-level category into three sub-categories: e.g., within proficient there is low-proficient, middle-proficient, and high-proficient. Each sub-category was constructed so that the range of included scale scores was less than the standard error of all scale scores within the broader proficiency category.

  Then the state convened a panel of experts to associate movements by students across those 12 sub-proficiency-level categories with the following descriptors (in reality, this was a "policy judgment" of what constituted "significant improvement," etc.):[3]
  - Significant decline = decrease of 3 or more sub-performance-level categories
  - Decline = decrease of 1 or 2 sub-performance-level categories
  - Maintain = no change in sub-performance-level category
  - Improvement = increase of 1 or 2 sub-performance-level categories
  - Significant improvement = increase of 3 or more sub-performance-level categories

  The state counted up students who fell into each of the above improvement categories according to the following table, which delineates between students who were previously proficient and those who were not:

[3] The one exception to this set of movement rules was if a student started out in the "advanced" proficiency category and exhibited a decline across the sub-levels of that "advanced" category, thus still remaining "advanced" – in this case the student was classified as "maintaining" and not as "decline" or "significant decline."

Previously Proficient?	Performance Level Change									
	Year t-1					Year t -2				
	SD	D	M	I	SI	SD	D	M	I	SI
Yes										
No										

For a school and subject, the state totaled the number of students with performance level change values separately by prior proficiency status and the applied the following weights to those counts:

Previously Proficient?	SD	D	M	I	SI
No	-2	-1	0	1	2
Yes	-2	-1	1	1	2

For example, the number of students who were previously proficient and made "significant declines" was multiplied by -2.

Then the state added up all of the weighted "performance level change" counts (across the two years in the table of Ns above) and divided that total by the sum of students with "performance level change" values across those same two years.

Finally, the state standardized the resultant average across schools of the same type (elementary/middle or high school) and this standardized value became the "change/improvement index value" for a given subject and school.

In cases where adjacent-year testing did not occur (i.e., for all calculations in high school grades as well as in subjects such as science, social studies, and writing), the state used a minimum of two and maximum of four years of prior, student-level standardized test scores (for subject X) to calculate yearly means. If the school had three or four years of data attached to it for subject X, the state fit a simple linear regression line to the mean scores and recorded the slope. If the school only had two years of prior scores with which to work, the state calculated the simple difference and took the result as the improvement value. The standardized value (across all comparable schools) of this change/slope became the "improvement index" for subject X for a given school.

Thus, the improvement index calculated for a given subject for a school could either be based on the standardized, weighted change in students' performance levels over three prior years or the standardized slope of average, standardized scale scores across two to four prior years.[4]

[4] Therefore, within the calculations for a school's "math context area index," the improvement component could be based on changes in students' performance levels – but for the same school, the improvement index component of the "science content area index" might be based on the slope approach.

- Gaps in student performance (weight = 25%): Once each student within a school had a standardized score (i.e., a z-score), the state arrayed all students attached to the school by z-score (within a subject area) for each of two prior years. Next the state took the average standardized scale score of the bottom 30 percent of students across those two years and subtracted from it the average standardized scale score of the top 30 percent of students across those two prior years to yield a subject-specific gap score. The resultant difference was standardized across comparable schools (i.e., elementary/middle or high school) and became the "achievement gap index" for subject X.

For each subject area, the state linearly combined scores for the above components according to weights in parentheses (0.5 for level achievement, 0.25 for growth, and 0.25 for gaps) to arrive at a "content area index" value, and standardized this value across schools.[5] For each school, the state took a simple linear combination of the standardized content-area indices (equally weighting each subject area) to arrive at the overall TTB index score.

If the school is a high school, its TTB index also included a subcomponent that was a function of prior four-year graduation rates: First, the state calculated a two-year average of a school's four-year graduation rate and standardized the resultant average across all schools. Second, the state computed the change (or slope) in four-year graduation rates based on a minimum of two years and a maximum of four years of historical graduation data, and standardized that slope across schools. The first part of this high school graduation index was weighted by two-thirds and the second part by one-third. The standardized value of their linear combination is the high school graduation index score. When calculating the overall TTB index score, the high school graduation index was weighted by 10 percent, with the remaining 90 percent evenly apportioned among the number of subjects for which a school had content area index values.

How does Michigan use the TTB ranking to identify Priority Schools?
Using the overall TTB index scores for all schools, the state ranked schools from highest to lowest. Next the state identified the subset of schools with TTB scores in the bottom 5 percent of this overall distribution. Federal guidance required that states identify the lowest performing 5 percent of Title I schools. Thus, once Michigan had identified the bottom 5 percent of all public schools, it counted the subset of those low-performing schools that were Title I (participating or eligible) and ensured that the resulting number was equal to or greater than 5 percent of the total number of Title I schools in the state in that year. In Michigan, the bottom 5 percent of all public schools included 5 percent of the stock of Title I public schools in the state.

How does Michigan use the TTB ranking to identify Focus Schools?
Michigan used one particular component of the overall TTB ranking to identify Focus schools: the achievement gap index. For each school, the state took a simple average of all available subject-area-specific achievement gap indices[6] to generate a composite gap index.

[5] In cases where performance level change information is not available for a school and in cases where the most recent year's proficiency rate is at or above 90%, the state omitted the "growth" part of the content area and weighted the remaining two components: level achievement and gaps.
[6] To identify the first cohort of Focus schools (in 2012), the state calculated the "average gap" by including all available subjects regardless of the number of FAY students with test scores (as long as the school had already met the criterion of having at least 30 FAY students in two subjects over two years). In all subsequent cohorts, the state only averaged subject-specific gap scores over subjects with at least 30 FAY students.

The state ranked schools by this composite gap index. Since federal waiver guidance required states to identify 10 percent of non-Priority, Title I schools with the largest achievement gaps as Focus, the state moved up the distribution of the composite gap index until it reached a value below which a number of non-Priority Title I schools equivalent to 10 percent of the population of Title I schools fell. Michigan labeled all schools, regardless of Title I status, below that cutoff as Focus schools.

Additional Changes to TTB Calculations for Subsequent Cohorts of Priority and Focus Schools

The most significant changes for this year were related to the identification of Focus schools. First, the state normalized students' scale scores before standardizing and examining within-school gaps between the average for the top 30 percent of students and the bottom 30 percent of students. Second, the state capped z-scores at 2 and -2. The intent of these adjustments was to reduce the capacity of outliers (especially in smaller schools) to disproportionately influence measures of gaps. Third, there was an "audit" function where a school that was flagged as Focus was removed if the bottom 30 percent of students in that school performed better than the state average in at least two subjects, and if that school's overall TTB ranking was above the 75th percentile.

Appendix B: References

Michigan Department of Education (2012a). "Top-to-Bottom Ranking, Priority, Focus, and Rewards Schools Identification Business Rules: 2011-2012." Technical Report.

Michigan Department of Education (2012b). "2012 Top-to-Bottom Ranking: Understanding How the Ranking is Calculated." Presentation.

Michigan Department of Education. "Top-to-Bottom School Rankings: Historical Ranking Information." Website: https://www.michigan.gov/mde/0,4615,7-140-22709_56562---00.html

Personal Communication with Venessa Keesler (March 18, 2016), Joseph Martineau (March 31, 2106) and Alex Schwarz (April 14, 2016)