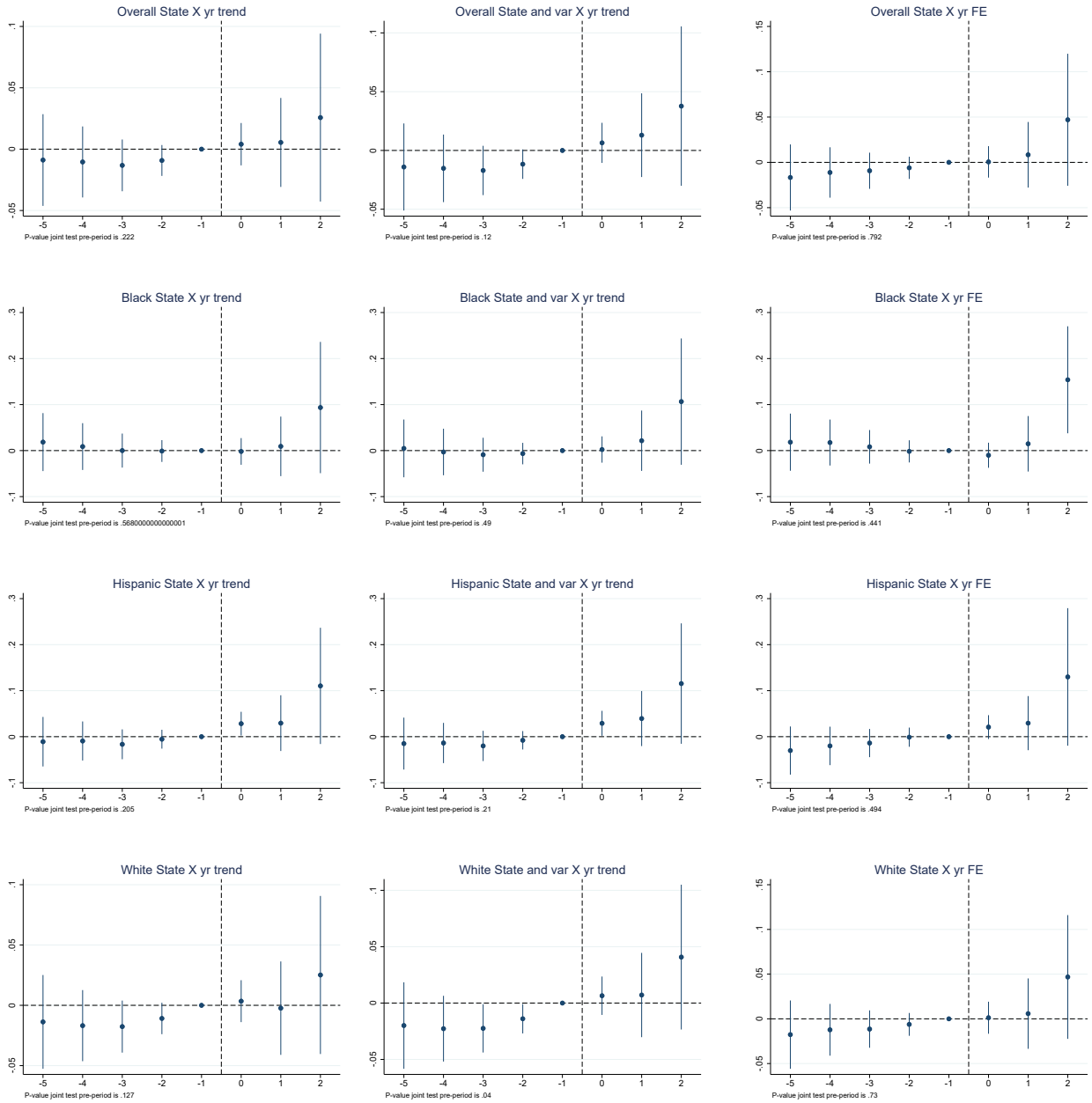
Online Appendix for Universal Access to Free School Meals and Student Achievement: Evidence from the Community Eligibility Provision

Krista Ruffini

Journal of Human Resources

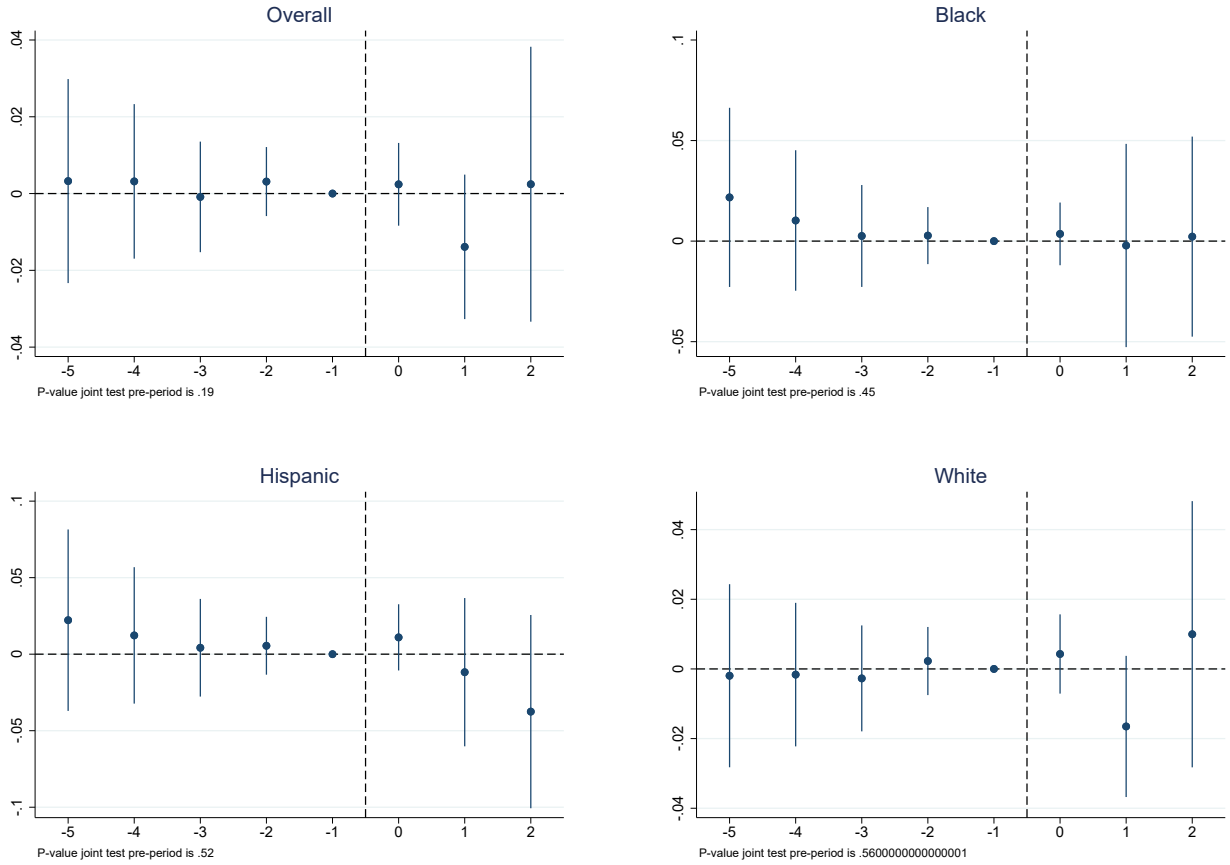
1 Appendix: Additional figures

Appendix Figure 1: Math Performance Event Study, Robustness, Exposed Districts



Notes: Figure presents results from event study framework in Equation 2. All specifications include controls for student demographics, the fraction of charter schools in a district, child poverty and unemployment rates, and measures of racial/ethnic segregation, year fixed effects, grade fixed effects, and district fixed effects. Left and center panels include state linear trends, center panel also includes linear trends in baseline covariates. Right panel includes state-by-year fixed effects. Bars denote 95 percent confidence intervals from robust standard errors clustered by district. Sample includes districts with a baseline FRP eligibility rate below 57.9 percent (the median among CEP-adopting districts). Notes below each panel present p-values from the joint test that pre-treatment coefficients equal to zero.

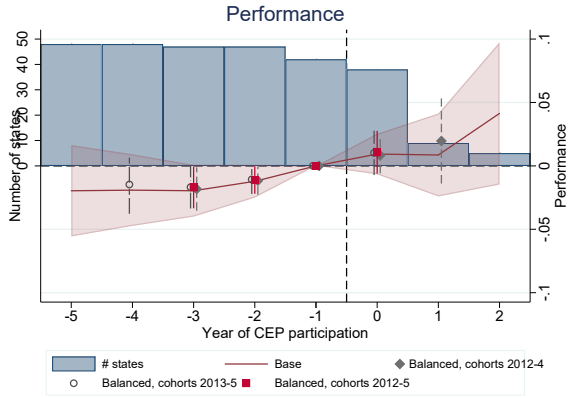
Appendix Figure 2: Reading Performance Event Study, Exposed Districts



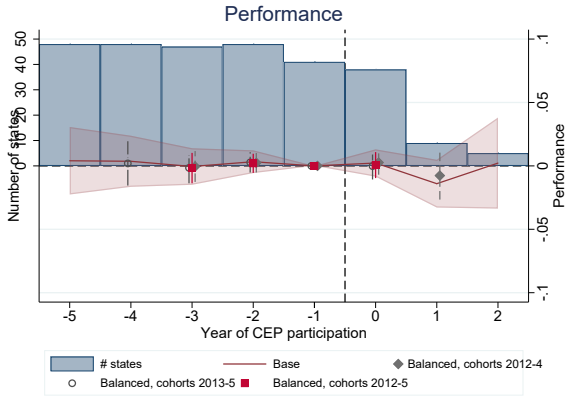
Notes: Figure presents results from the (district-level) event study framework in Equation 2. All specifications include controls for student demographics, the fraction of charter schools in a district, child poverty and unemployment rates, and measures of racial/ethnic segregation, year fixed effects, grade fixed effects, and district fixed effects. Bars denote 95 percent confidence intervals from robust standard errors clustered by district. Sample includes districts with a baseline FRP eligibility rate below 57.9 percent (the median among CEP-adopting districts). Notes below each panel present p-values from the joint test that pre-treatment coefficients equal to zero.

Appendix Figure 3: Overall Performance: Balanced and Unbalanced Event Studies

(a) Math

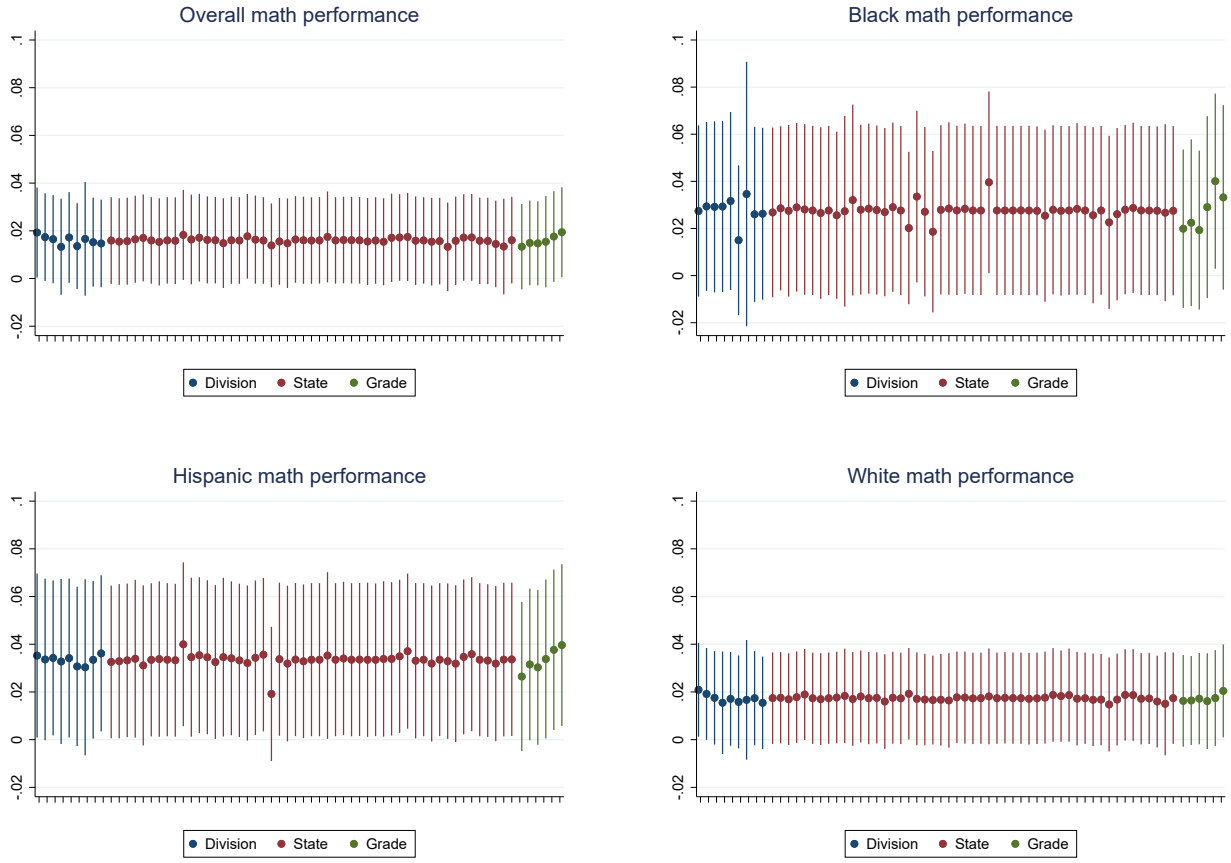


(b) Reading



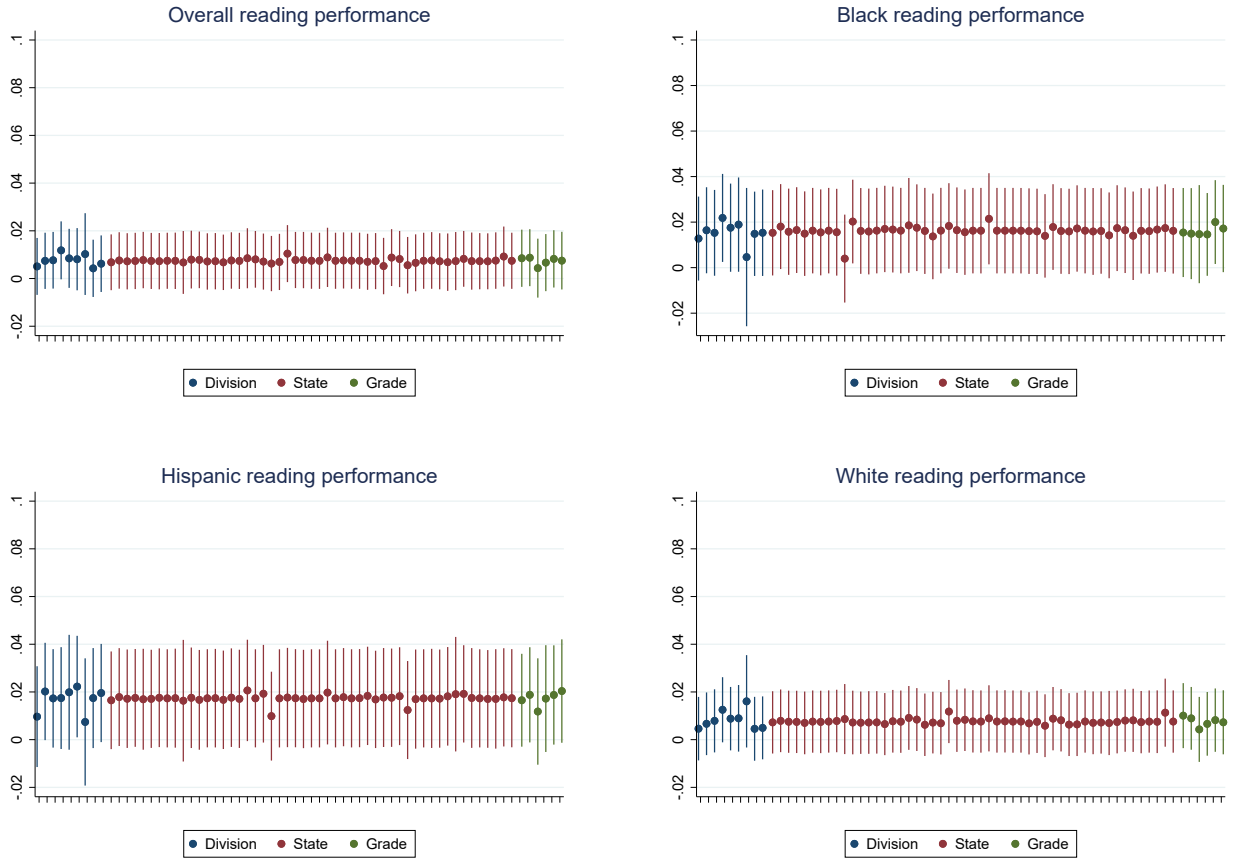
Notes: Figure summarizes the number of states contributing to each event year in the unbalanced panel (blue bars), and presents results from event study framework in Equation 2, with event years defined as year relative to CEP implementation for both unbalanced (maroon line) and three balanced subpanels. The gray diamonds show the balanced panel among districts that first adopted CEP between 2012 and 2014; the open gray circles show the 2013-2015 cohorts; and the bright red squares show the balanced event study for districts that adopted within the 2012 through 2015 period. All specifications include controls for student demographics, the fraction of charter schools in a district, child poverty and unemployment rates, and measures of racial/ethnic segregation, year fixed effects, grade fixed effects, and district fixed effects. 95 percent confidence intervals from robust standard errors clustered by district. Sample includes districts with a baseline FRP eligibility rate below 57.9 percent (the median among CEP-adopting districts).

Appendix Figure 4: Math Performance: Drop Division, State, Grade



Notes: Figure plots coefficients and confidence intervals from the specifications in Table 6 for district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017, but dropping a single Census Division (blue), state (red), or grade (green). All omitted areas and grades are in ascending order (e.g.: the far-left point is Census Division 1, Alabama, or grade 3, the far-right point is Census Division 9, Wyoming, or grade 8. This figure indicates that results are not driven by the experiences of a single state or geographic area. Consistent with Table 7, math performance gains tend to be larger for younger grades. All specifications include district, grade, and year fixed effects, as well as student racial/ethnic composition and segregation, student-teacher ratios, percent of students attending a charter school, child poverty rates and county unemployment rates. Bars denote 95 percent confidence intervals from robust standard errors clustered by district.

Appendix Figure 5: Reading Performance: Drop Division, State, Grade



Notes: Figure plots coefficients and confidence intervals from the specifications in Table 6 for district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017, but dropping a single Census Division (blue), state (red), or grade (green). All omitted areas and grades are in ascending order (e.g.: the far-left point is Census Division 1, Alabama, or grade 3, the far-right point is Census Division 9, Wyoming, or grade 8. This figure indicates that results are not driven by the experiences of a single state or geographic area. Consistent with Table 7, math performance gains tend to be larger for younger grades. All specifications include district, grade, and year fixed effects, as well as student racial/ethnic composition and segregation, student-teacher ratios, percent of students attending a charter school, child poverty rates and county unemployment rates. Bars denote 95 percent confidence intervals from robust standard errors clustered by district.

2 Appendix: Additional tables

Appendix Table 1: Effect of CEP on Meal Consumption: Parametric Event Study

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Breakfast				Lunch		
	All	All	Exposed	Exposed	All	All	Exposed	Exposed
Event year	5.982	4.981	2.601**	1.979	4.808	3.703	0.648	-4.487
	(3.460)	(3.419)	(0.970)	(5.410)	(2.966)	(2.813)	(1.229)	(2.635)
Post	9.983***	9.767***	13.330***	12.379***	10.245***	10.329***	10.433***	10.101***
	(1.719)	(1.608)	(1.838)	(1.749)	(0.851)	(0.828)	(1.036)	(1.100)
Event year X post	0.078	0.206	1.839	0.769	1.701	1.917	3.152***	3.231***
	(0.996)	(1.056)	(1.544)	(1.510)	(0.978)	(0.984)	(0.780)	(0.886)
StateXyear trends	X	X	X	X	X	X	X	X
Baseline var trends		X		X		X		X
Observations	14248	14248	6003	6003	14269	14269	6013	6013

Notes: Table presents unweighted results from meal count data collected from state Department of Educations for six of the eleven states that adopted CEP before 2015: Georgia, Illinois, Kentucky, New York, Maryland, and West Virginia. Data availability varies by state, but spans 2009-2016. All specifications include controls for student demographics, the fraction of charter schools in a district, child poverty and unemployment rates, and measures of racial/ethnic segregation, as well as year and school fixed effects. Even-numbered columns also include state-specific linear trends and trends in baseline variables. Robust standard errors clustered by district. Columns (1-2) and (5-6) ("all") include all observations that adopted CEP between 2012 and 2017; columns (3-4) and (7-8) ("exposed") restrict the sample to observations in districts with a baseline FRP eligibility rate below 57.9 percent (the median among CEP-adopting districts). $\beta_{ey} = \beta_{ey*post}$ presents p-value from a hypothesis test that the pre-CEP linear trends equals the trend after CEP adoption. Robust standard errors clustered by district. See text and data appendix for details. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.10$.

Appendix Table 2: Effect of CEP on Meal Consumption: Linear Trends by State and Baseline Variables

	(1)	(2)	(3)	(4)	(5)	(6)
	School per-student breakfast		School per-student lunch		Log per student nutrit asst	
	All	Exposed	All	Exposed	All	Exposed
CEP	12.102***	12.520***	12.371***	12.129***	0.074***	0.082***
	(2.167)	(2.754)	(1.259)	(1.415)	(0.009)	(0.012)
Observations	18762	12077	20030	13193	128145	64105
Basline DV mean	52.57	49.16	111.9	104.3	0.400	0.327
Pct change	0.230	0.255	0.111	0.116		
StateXyear trends	X	X	X	X	X	X
Baseline trends	X	X	X	X	X	X
Level	School	School	School	School	District	District

Notes: Table presents unweighted results from estimating Equation 1 at the school level (columns (1) through (4)) with meal count data collected from state Department of Educations for six of the eleven states that adopted CEP before 2015: Georgia, Illinois, Kentucky, New York, Maryland, and West Virginia. Data availability varies by state, but spans 2009-2016. Columns 5 and 6 presents federal nutritional assistance dollars, reported in the Annual Survey of School System Finances. All specifications include controls for student demographics, the fraction of charter schools in a district, child poverty and unemployment rates, and measures of racial/ethnic segregation, as well as year and school fixed effects. Robust standard errors clustered by district. Odd-numbered columns ("all") include all observations that adopted CEP between 2012 and 2017; even-numbered columns ("exposed") restrict the sample to observations in districts with a baseline FRP eligibility rate below 57.9 percent (the median among CEP-adopting districts). Robust standard errors clustered by district. See text and data appendix for details. *** $= p < 0.01$, ** $= p < 0.05$, * $= p < 0.10$.

Appendix Table 3: Federal Nutritional Assistance ($1,000s) and Overall Student Performance

	(1) Math	(2) Math	(3) Reading	(4) Reading
Per-student fed nutr. asst	0.163	0.512*	-0.142	0.235
	(0.212)	(0.296)	(0.142)	(0.206)
Observations	59465	31423	62174	32968
Sample	All	Exposed	All	Exposed
Baseline DV mean (level)	-.247	-.118	-.232	-.095
Change in nutritional asst	0.100	0.094	0.099	0.093
F stat 1st stage	184.856	106.867	156.994	105.841

Notes: Table presents 2SLS regression results where the change in per-student federal nutritional assistance is instrumented by CEP participation. "Exposed" districts are district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017; treatment districts are districts in which at least one school adopts CEP by 2015. "All" districts include all district-grade observations that participated in CEP at any point by 2017. All specifications include district, grade, and year fixed effects, as well as student racial/ethnic composition and segregation, percent of students attending a charter school, child poverty rates and county unemployment rates. All specifications are weighted least squares, with weights equal to the squared inverse of the standard error of the district-grade performance metric. Robust standard errors clustered by district. See text and data appendix for details. *** $= p < 0.01$, ** $= p < 0.05$, * $= p < 0.10$.

Appendix Table 4: Predicted Performance from Changes in Racial/Ethnic Composition

	(1)	(2)	(3)	(4)
	$\widehat{Overall}$	$\widehat{Black}$	$\widehat{Hispanic}$	$\widehat{White}$
	Panel A: Math performance			
CEP	-0.001	-0.002**	0.001	0.001
	(0.001)	(0.001)	(0.003)	(0.001)
Observations	32694	11658	12698	29325
Baseline FRP	0.454	0.457	0.438	0.458
Baseline DV mean	0.0773	-0.418	-0.225	0.214
	Panel B: Reading performance			
CEP	-0.000	-0.002**	-0.001	0.002
	(0.001)	(0.001)	(0.002)	(0.001)
Observations	34344	12185	13256	30581
Baseline FRP	0.453	0.457	0.436	0.458
Baseline DV mean	0.0745	-0.421	-0.226	0.213
Area and district controls	X	X	X	X
Sample	Exposed	Exposed	Exposed	Exposed

Notes: Table presents weighted least squares regression results from the specification in Equation 1 for district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017; treatment districts are districts in which at least one school adopts CEP by 2015. Race/ethnic proficiency scores available for cells with at least 20 students. All specifications include district, grade, and year fixed effects, student-teacher ratios, percent of students attending a charter school, child poverty rates and county unemployment rates. Dependent variable is defined as predicted values from a regression interacting each grade with the share of students of each racial/ethnic group in a district and CEP schools within a district, as well as the dissimilarity index for each racial/ethnic group. Robust standard errors clustered by district. See text and data appendix for details. *** $= p < 0.01$, ** $= p < 0.05$, * $= p < 0.10$.

Appendix Table 5: Effect of CEP on Math Performance: High-Exposure Districts Sample, Alternative Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: All			
CEP	0.015	0.011	0.008	0.013	0.002	0.016*	0.018**
	(0.009)	(0.009)	(0.007)	(0.009)	(0.010)	(0.008)	(0.008)
Observations	32694	32607	26301	32694	32694	32694	32645
Baseline FRP	0.454	0.454	0.455	0.454	0.454	0.454	0.454
Baseline DV mean	-0.121	-0.121	-0.116	-0.121	-0.121	-0.121	-0.121
				Panel B: Black			
CEP	0.024	0.011	0.010	0.022	-0.002	0.001	0.006
	(0.018)	(0.016)	(0.014)	(0.020)	(0.021)	(0.014)	(0.014)
Observations	11658	11658	8996	11658	11658	11658	11658
Baseline FRP	0.457	0.457	0.459	0.457	0.457	0.457	0.457
Baseline DV mean	-0.502	-0.502	-0.500	-0.502	-0.502	-0.502	-0.502
				Panel C: Hispanic			
CEP	0.031**	0.029**	0.029***	0.027*	0.026	0.025*	0.029**
	(0.016)	(0.014)	(0.011)	(0.016)	(0.018)	(0.013)	(0.013)
Observations	12698	12679	9582	12698	12698	12698	12679
Baseline FRP	0.438	0.438	0.437	0.438	0.438	0.438	0.438
Baseline DV mean	-0.315	-0.315	-0.305	-0.315	-0.315	-0.315	-0.315
				Panel D: White			
CEP	0.017*	0.010	0.006	0.016*	-0.001	0.019**	0.021**
	(0.010)	(0.009)	(0.008)	(0.010)	(0.011)	(0.009)	(0.009)
Observations	29325	29272	23324	29325	29325	29325	29293
Baseline FRP	0.458	0.458	0.459	0.458	0.458	0.458	0.458
Baseline DV mean	0.011	0.011	0.015	0.011	0.011	0.011	0.011
Treatment defn	Binary	Binary	Binary	Binary	Pct	Binary	Binary
Resource variables	X						
State X year trends		X	X				
Baseline trends		X	X				
Lagged performance			X				
State X year FE				X			
Weights	WLS	WLS	WLS	WLS	WLS	District	Log enroll

Notes: Table presents regression results from the specification in Equation 1 for district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017; treatment districts are districts in which at least one school adopts CEP by 2015. Race/ethnic proficiency scores available for cells with at least 20 students. All specifications include district, grade, and year fixed effects, as well as student racial/ethnic composition and segregation, student-teacher ratios, percent of students attending a charter school, child poverty rates and county unemployment rates. "Resource variables" include per-pupil total and instructional expenditures; "baseline trends" includes linear trends for baseline values of all control variables. Columns (1-4) present weighted least squares regressions with additional controls; column (5) presents unweighted results; column (6) weights each observation by the log number of students in each racial/ethnic group between 2009 and 2011. Robust standard errors clustered by district. See text and data appendix for details. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.10$.

Appendix Table 6: Effect of CEP on Reading Performance: High-Exposure Districts Sample, Alternative Specifications

	(1)	(2)	(3)	(4)		(6)	(7)
				Panel A: All			
CEP	0.007	0.005	0.001	-0.001	-0.002	0.007	0.008
	(0.006)	(0.005)	(0.004)	(0.005)	(0.007)	(0.007)	(0.006)
Observations	34344	34250	28329	34344	34344	34344	34291
Baseline FRP	0.453	0.453	0.454	0.453	0.453	0.453	0.453
Baseline DV mean	-0.104	-0.104	-0.101	-0.104	-0.104	-0.104	-0.104
				Panel B: Black			
CEP	0.014	0.008	0.001	0.006	-0.019	-0.007	-0.002
	(0.009)	(0.007)	(0.007)	(0.009)	(0.017)	(0.011)	(0.011)
Observations	12185	12185	9504	12185	12185	12185	12185
Baseline FRP	0.457	0.457	0.458	0.457	0.457	0.457	0.457
Baseline DV mean	-0.441	-0.441	-0.441	-0.441	-0.441	-0.441	-0.441
				Panel C: Hispanic			
CEP	0.016	0.018**	0.005	0.011	0.014	-0.000	0.006
	(0.010)	(0.009)	(0.009)	(0.010)	(0.015)	(0.011)	(0.011)
Observations	13256	13234	10110	13256	13256	13256	13236
Baseline FRP	0.436	0.436	0.434	0.436	0.436	0.436	0.436
Baseline DV mean	-0.391	-0.391	-0.385	-0.391	-0.391	-0.391	-0.391
				Panel D: White			
CEP	0.007	0.004	-0.000	-0.001	-0.004	0.010	0.011
	(0.007)	(0.006)	(0.005)	(0.006)	(0.008)	(0.007)	(0.007)
Observations	30581	30530	24789	30581	30581	30581	30550
Baseline FRP	0.458	0.458	0.459	0.458	0.458	0.458	0.458
Baseline DV mean	0.051	0.051	0.054	0.051	0.051	0.051	0.051
Treatment defn	Binary	Binary	Binary	Binary	Pct	Binary	Binary
Resource variables	X						
State X year trends		X	X				
Baseline trends		X	X				
Lagged performance			X				
State X year FE				X			
Weights	WLS	WLS	WLS	WLS	WLS	District	Log enroll

Notes: Table presents regression results from the specification in Equation 1 for district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017; treatment districts are districts in which at least one school adopts CEP by 2015. Race/ethnic proficiency scores available for cells with at least 20 students. All specifications include district, grade, and year fixed effects, as well as student racial/ethnic composition and segregation, student-teacher ratios, percent of students attending a charter school, child poverty rates and county unemployment rates. "Resource variables" include per-pupil total and instructional expenditures; "baseline trends" includes linear trends for baseline values of all control variables. Columns (1-4) present weighted least squares regressions with additional controls; column (5) presents unweighted results; column (6) weights each observation by the log number of students in each racial/ethnic group between 2009 and 2011. Robust standard errors clustered by district. See text and data appendix for details. *** $= p < 0.01$, ** $= p < 0.05$, * $= p < 0.10$.

Appendix Table 7: Effect of CEP on Math Performance: High-Exposure Districts Sample, Alternative Samples

	(1)	(2)	(3)
	Panel A: All		
CEP	0.014	0.015	0.005
	(0.016)	(0.011)	(0.016)
Observations	14835	22162	12948
Baseline FRP	0.468	0.458	0.455
Baseline DV mean	-0.168	-0.098	-0.126
	Panel B: Black		
CEP	-0.024	0.021	0.017
	(0.035)	(0.023)	(0.031)
Observations	2734	6636	5228
Baseline FRP	0.502	0.465	0.455
Baseline DV mean	-0.546	-0.487	-0.495
	Panel C: Hispanic		
CEP	0.007	0.031	0.010
	(0.037)	(0.019)	(0.031)
Observations	2956	6552	6146
Baseline FRP	0.453	0.433	0.438
Baseline DV mean	-0.365	-0.273	-0.289
	Panel D: White		
CEP	0.021	0.017	0.001
	(0.017)	(0.012)	(0.017)
Observations	12954	18949	11350
Baseline FRP	0.474	0.462	0.462
Baseline DV mean	-0.074	0.040	0.024
Sample	Full dist	Balanced	Adopt 1st yr

Notes: Table presents weighted least squares results from the specification in Equation 1 for district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017. Race/ethnic proficiency scores available for cells with at least 20 students. 'All specifications include district, grade, and year fixed effects, as well as student racial/ethnic composition and segregation, student-teacher ratios, percent of students attending a charter school, child poverty rates and county unemployment rates. Column (1) restricts to district-grade observations where every school serving grade g participates in CEP upon CEP adoption. Column (2) limits the sample to district-grade observations with a valid performance score each year. Column (3) limits the sample to districts that participated in CEP the first year their state became eligible. Robust standard errors clustered by district. See text and data appendix for details. *** $= p < 0.01$, ** $= p < 0.05$, * $= p < 0.10$.

Appendix Table 8: Effect of CEP on Reading Performance: High-Exposure Districts Sample, Alternative Samples

	(1)	(2)	(3)
	Panel A: All		
CEP	0.021*	-0.000	-0.021***
	(0.013)	(0.006)	(0.008)
Observations	15373	26999	13720
Baseline FRP	0.467	0.457	0.454
Baseline DV mean	-0.144	-0.096	-0.142
	Panel B: Black		
CEP	-0.007	0.005	0.003
	(0.027)	(0.010)	(0.012)
Observations	2782	7504	5397
Baseline FRP	0.501	0.462	0.455
Baseline DV mean	-0.472	-0.445	-0.462
	Panel C: Hispanic		
CEP	0.004	0.006	0.002
	(0.032)	(0.010)	(0.011)
Observations	3037	7028	6432
Baseline FRP	0.448	0.427	0.435
Baseline DV mean	-0.477	-0.388	-0.413
	Panel D: White		
CEP	0.024*	0.001	-0.028***
	(0.014)	(0.007)	(0.009)
Observations	13295	22302	11861
Baseline FRP	0.474	0.462	0.462
Baseline DV mean	-0.025	0.053	0.046
Sample	Full dist participation	Balanced panel	Adopt 1st yr eligibility

Notes: Table presents weighted least squares results from the specification in Equation 1 for district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017. Race/ethnic proficiency scores available for cells with at least 20 students. 'All specifications include district, grade, and year fixed effects, as well as student racial/ethnic composition and segregation, student-teacher ratios, percent of students attending a charter school, child poverty rates and county unemployment rates. Column (1) restricts to district-grade observations where every school serving grade g participates in CEP upon CEP adoption. Column (2) limits the sample to district-grade observations with a valid performance score each year. Column (3) limits the sample to districts that participated in CEP the first year their state became eligible. Robust standard errors clustered by district. See text and data appendix for details. *** $= p < 0.01$, ** $= p < 0.05$, * $= p < 0.10$.

Appendix Table 9: Effects of CEP on Reading Performance, Exposure Distribution

	(1)	(2)	(3)	(4)	(5)
Baseline FRP eligible	$\leq 40\%$	$\leq 50\%$	$\leq 60\%$	$\leq 70\%$	$\leq 80\%$
	Panel A: Overall performance				
CEP	0.012	0.006	0.006	0.001	-0.003
	(0.013)	(0.007)	(0.006)	(0.006)	(0.005)
Observations	8054	21276	37599	51906	61616
Average baseline FRP	0.312	0.401	0.465	0.515	0.551
Baseline DV mean	-0.034	-0.065	-0.116	-0.175	-0.220
	Panel B: Black performance				
CEP	0.030	0.015	0.010	-0.003	-0.006
	(0.028)	(0.012)	(0.009)	(0.010)	(0.009)
Observations	2834	7631	13590	21013	27147
Average baseline FRP	0.325	0.407	0.471	0.534	0.582
Baseline DV mean	-0.383	-0.415	-0.450	-0.488	-0.510
	Panel C: Hispanic performance				
CEP	0.033	0.016	0.013	0.003	0.004
	(0.023)	(0.013)	(0.010)	(0.009)	(0.008)
Observations	3976	8605	14499	19961	23970
Average baseline FRP	0.294	0.380	0.449	0.503	0.543
Baseline DV mean	-0.358	-0.363	-0.394	-0.422	-0.445
	Panel D: White performance				
CEP	0.011	0.004	0.008	0.004	0.002
	(0.014)	(0.008)	(0.006)	(0.006)	(0.005)
Observations	6703	19017	33385	44762	51077
Average baseline FRP	0.326	0.410	0.469	0.514	0.542
Baseline DV mean	0.148	0.079	0.046	0.023	0.006
Percentile baseline FRP distribution	11.700	31.000	54.800	75.500	89.600

Notes: Table presents weighted least squares regression results from Equation 1 for all district-grade observations in which any school serving grade g participated in CEP by 2017 based on the baseline (2009-2011) share of students FRP eligible under the traditional formula. CEP equals one if any school serving grade g in district d participated in CEP by year t. Race/ethnic proficiency scores available for cells with at least 20 students. "Average baseline FRP" indicates average baseline (2009-2011) eligibility rates. "Percentile baseline FRP distribution" displays the share of districts with baseline eligibility $\leq x\%$. All specifications include district, grade, and year fixed effects, as well as student racial/ethnic composition and segregation, student-teacher ratios, percent of students attending a charter school, child poverty rates and county unemployment rates. Robust standard errors clustered by district. See text and data appendix for details. *** $= p < 0.01$, ** $= p < 0.05$, * $= p < 0.10$.

3 Data appendix

CEP participation data

School-level CEP participation information comes from state educational agencies for the 2012 through 2014 pilot years. Data for the 2015 through 2017 academic years are provided for most states in the Common Core of Data, and for the remaining states by FRAC and CBPP. States with CEP participation in the CCD are identified by a unique NCES-issued district identifier. For states with incomplete CEP participation in the CCD (Illinois, West Virginia, the District of Columbia, North Carolina, Utah, Vermont, and Wyoming), I standardize all school and district names and match to the Common Core of Data (CCD) based on state, district, and school name. This procedure matches approximately 93 percent for all public elementary and middle schools.[1]

The CCD also provides grade-level enrollment information necessary to collapse the school-level CEP participation data to the grade-district level. The main analyses define a binary treatment variable equal to one if any school in district d serving grade g at time t had implemented CEP. For example, if in 2015, a district has a CEP school with some third graders, but no eighth graders, attending a school offering CEP, third graders are in the district are considered "treated," but eighth graders are considered "untreated." In practice, conditional on having *any* district participation in grade 3-8, 72 percent of districts have participation at each grade level, with slightly higher participation rates in younger grades. In robustness checks (Appendix Tables 5 and 6), I calculate a continuous measure of CEP participation, measured as the fraction of students in district d and grade g attending a CEP school in year t, using school-level enrollment data from the CCD.

Meal count data

No existing data set provides a consistent measure of school meal receipt across states or over time. Schools and states report this information to USDA for federal reimbursement; however, the recorded meal count measures and the duration of maintained records varies by state. Between September 2016 and May 2017, research assistants contacted state staff in each of the eleven pilot states. Of these states, six provided school-level information on the number of breakfasts and lunches served in each school, and three provided this information by payment status. Following the string matching procedure described above, I match school-level meal receipt data to school-level CEP participation and demographic information from the CCD. In order to construct a comparable measure of meal consumption across states, I calculate the number of breakfasts and lunches per student, where student enrollment is provided in the CCD.

District performance data

I merge the district-level CEP data to district-grade-year performance data from the Stanford Education Data Archive (SEDA, version 2.1) using NCES district-level identifiers, with charter schools placed in the district in which they are geographically located. The SEDA data is a novel dataset that is unique in its ability to compare achievement across states

[1] An earlier version of this paper matched schools using fuzzy string match to minimize the Levenshtein distance. These approaches yield almost identical empirical results; the current approach is more accurate.

and over time at the substate (e.g.: district) level by combining information from the state-level National Assessment of Educational Progress (NAEP) results and restricted-use state-proficiency data. Reardon et al. (2018) detail the data construction, variable definitions, and underlying assumptions. I highlight several key points here regarding measurement and sample restrictions, and discuss the feasibility of alternative performance measures.

Performance measurement:

As required by federal legislation, each state administers reading and math examinations to every student in grades 3-8. The number of students scoring "proficient" are reported at the school level. Each state has its own "proficiency" standard and these definitions vary both across states and over time.[2]

The SEDA framework first applies a heteroskedastic ordered probit (HETOP) model (homoskedastic ordered probit for states with only two proficiency categories) to the proficiency categories reported to the Department of Education to estimate the mean and standard deviation for each state-subject-grade-year at the district level, as well as the corresponding standard errors. Intuitively, this step transforms categorical proficiency measures to a continuous measure. These means and standard deviations are then standardized by the state-subject-grade-year distribution to have a mean of zero and a standard deviation of one. Scores for each race/ethnic subgroup are based on the overall cutpoints so that the performance scale for each subgroup matches the overall distribution.

In order to facilitate comparisons across states and over time, the SEDA data adjusts these state-grade-year-subject estimates with information from state-level National Assessment of Educational Progress (NAEP) results. NAEP is an examination that is administered biennially to fourth and eighth graders in a sample of districts, and is designed to yield measures of math and reading achievement that are comparable across states and over time (see Reardon et al. (2017) for a detailed discussion). Each test wave, NAEP is designed to be nationally representative (and representative at the state level for math and reading), but only a sample of schools and students are chosen to participate, which limits its ability to track student performance at the school- or district-level over time.[3]

The SEDA framework takes the state-level NAEP data and first interpolates and extrapolates each state mean and standard deviation to years and grade levels not covered in the NAEP (e.g.: even-numbered years and grades 3 and 5-7). It then places the district-grade-subject-year continuous proficiency measures from the state-assessment data on the cross-state NAEP scale. A technical discussion is provided in Reardon et al. (2018) (Equations 6.2 and 6.3). At an intuitive level, districts that perform well on their state's assessment are placed on the SEDA scale high relative to their state's NAEP measure, and districts in states that score higher on the NAEP assessment also place higher in the SEDA performance distribution.

[2]A substantial change occurred in the 2014-2015 school year, when the 34 states with waivers from the No Child Left Behind proficiency standards were required to give "high-quality assessments" of "college- and career-ready standards." These examinations (e.g. PARCC and Smarter Balance) tended to have more stringent proficiency requirements than the earlier state examinations Education Week (2014) and resulted in a sharp drop in the number of students achieving proficiency.

[3]District-level results are available for 21 of the largest districts that participate in the Trial Urban District Assessment (TUDA) pilot.

Finally, the NAEP-proficiency estimates are standardized to the national student-level NAEP score distributions (for years and grades the NAEP was administered) or the interpolated/ extrapolated NAEP score distributions (for observations the NAEP was not administered). All results in this paper use the "cohort standardized (cs)" scale, which provides a measure of performance that is comparable over time at the district-grade-subject level ((Reardon et al., 2018) Equation 7.1). Under this measure, treatment effects are provided in effect sizes.

Following Reardon et al. (2018), all models are estimated using weighted least squares, with weights inversely proportional to the estimated variance of the performance metric. These weights, detailed in Reardon et al. (2019) account for the NAEP interpolation and extrapolations, as well as other sources of linking error.

Excluded observations

Performance measures are not included for district-years that administer locally-selected examinations (this mainly affects middle schools in California, Virginia, and Texas in some years), district-grade-years with participation rates below 95 percent, instances where district-grade-year information was not reported to the Department of Education, cases of identified data errors in state proficiency data, or cases with estimated standard errors greater than the state-standardized metric. Finally, random noise is added to each estimate and district-subject-grade-year cells from fewer than 20 assessments are suppressed for confidentiality issues. As this random noise introduces classical measurement error, it will slightly attenuate the reported results. Data Appendix Table 1 lists the state-subject-grades that are not available; Reardon et al. (2018) further describes the rationale for exclusion.

The final dataset has approximately 66,000 (math) to 69,000 (reading) district-year-grade observations for locations that participated in CEP through 2017, with subgroup sample sizes ranging from 25,000 (Hispanic math performance) to 53,000 (white reading performance). My main results focus on the 32,000-34,000 district-grade-year observations with relatively low district eligibility for the free meals program before CEP (those where less than 57.9 percent of students were eligible for free meals between 2009-2011).

In order to verify that CEP adoption is not associated with whether a valid score exists for each district-grade-subject-year, I create a balanced panel of district-grade-subjects and create a binary outcome variable equal to one if the district-grade-subject-subgroup-year is available in the SEDA data. Data Appendix Table 2 presents the results for math and reading and shows for overall, Hispanic, and white subgroups, there is no economically or statistically significant relationship between CEP implementation and the availability of SEDA data. For black performance, CEP adoption is associated with a 2 percentage point lower probability a district-subject-grade-year appears in the SEDA data. However, the results from the balanced sample of district-grades in Column (2) of Appendix Tables 7 and 8 indicate that differential attrition from the SEDA data is not driving the main results.

Data Appendix Table 1: Grade-subject-years not available in the SEDA data

	Math							Reading						
	2009	2010	2011	2012	2013	2014	2015	2009	2010	2011	2012	2013	2014	2015
Arkansas	8	8					8							
California	7-8	7-8	7-8	7-8	7-8	3-8							3-8	
Colorado	3-8	3-8	3-8					3-8	3-8	3-8				
Connecticut						3-8							3-8	
Delaware							8							8
Florida						3-8								
Idaho						3-8							3-8	
Kansas						3-8							3-8	
Maine							6-8							6-7
Maryland						3-4, 6-7							3-4, 6-7	
Missouri					8	8	8							
Montana						3-8	3-8						3-8	3-8
Nebraska	3-8	3-8						3-8						
Nevada						3-8	3-8						3-8	3-8
New Hampshire							8							8
New Jersey							3-8							3-8
New York						6-8	3-8						6-8	3-8
North Dakota							3-8							3-8
Ohio							8							
Oklahoma				8	8									
Oregon						3-8							3, 7-8	
Rhode Island							6-8							5-8
South Dakota						3-8								3-8
Tennessee							8							
Texas				7-8	7-8	7-8	7-8							
Utah	8	8	8	8	8									
Virginia	5-8	5-8	5-8	5-8	5-8	5-8	5-8							
Washington						3-8	3-8						3-8	3-8
West Virginia				3-8		3-7					3-8			
Wyoming		3-8			3-8	3, 7-8					3-8		3-8	

Source: Reardon et al. (2018) Table A1

Data Appendix Table 2: SEDA availability and CEP adoption

	(1)	(2)	(3)	(4)
	SEDA exists			
	Overall	Black	Hispanic	White
	Panel A: Math performance			
CEP	-0.002	-0.021***	0.001	-0.000
	(0.004)	(0.008)	(0.008)	(0.008)
Observations	37114	37114	37114	37114
Baseline FRP	0.454	0.454	0.454	0.454
Baseline DV mean	0.949	0.334	0.347	0.840
	Panel B: Reading performance			
CEP	0.001	-0.018**	0.006	0.011
	(0.004)	(0.008)	(0.009)	(0.008)
Observations	37114	37114	37114	37114
Baseline FRP	0.453	0.453	0.453	0.453
Baseline DV mean	0.967	0.342	0.356	0.853
Sample	Exposed	Exposed	Exposed	Exposed

Table presents weighted least squares regression results from the specification in Equation 1 for district-grade observations with a baseline FRP eligibility share below 57.9 percent (the baseline median among CEP districts) in which any school serving grade g participated in CEP by 2017. Dependent variable equals one if a measure of student performance exists for the district-grade-subject-year in the SEDA data. Robust standard errors clustered by district. See text and data appendix for details. *** $= p < 0.01$, ** $= p < 0.05$, * $= p < 0.10$.

Other performance measures

The SEDA data is only data source that provide substate measures of academic performance that are consistently available at the substate level and are comparable over time. Here I discuss other measures of student performance that are less well-suited to this analysis.

First, the NAEP data are not administered to the same sample of schools each wave, limiting the potential for these data to draw comparisons within a district over time.

Second, as described in Reardon et al. (2018), state proficiency examinations change over time, both in the assessment battery and the proficiency definition. Therefore, comparing outcomes across states using state proficiency data is problematic, even in models that include state and year fixed effects. Of particular concern for this analysis, the pilot and national implementation CEP period coincides with three important changes to state proficiency examinations:

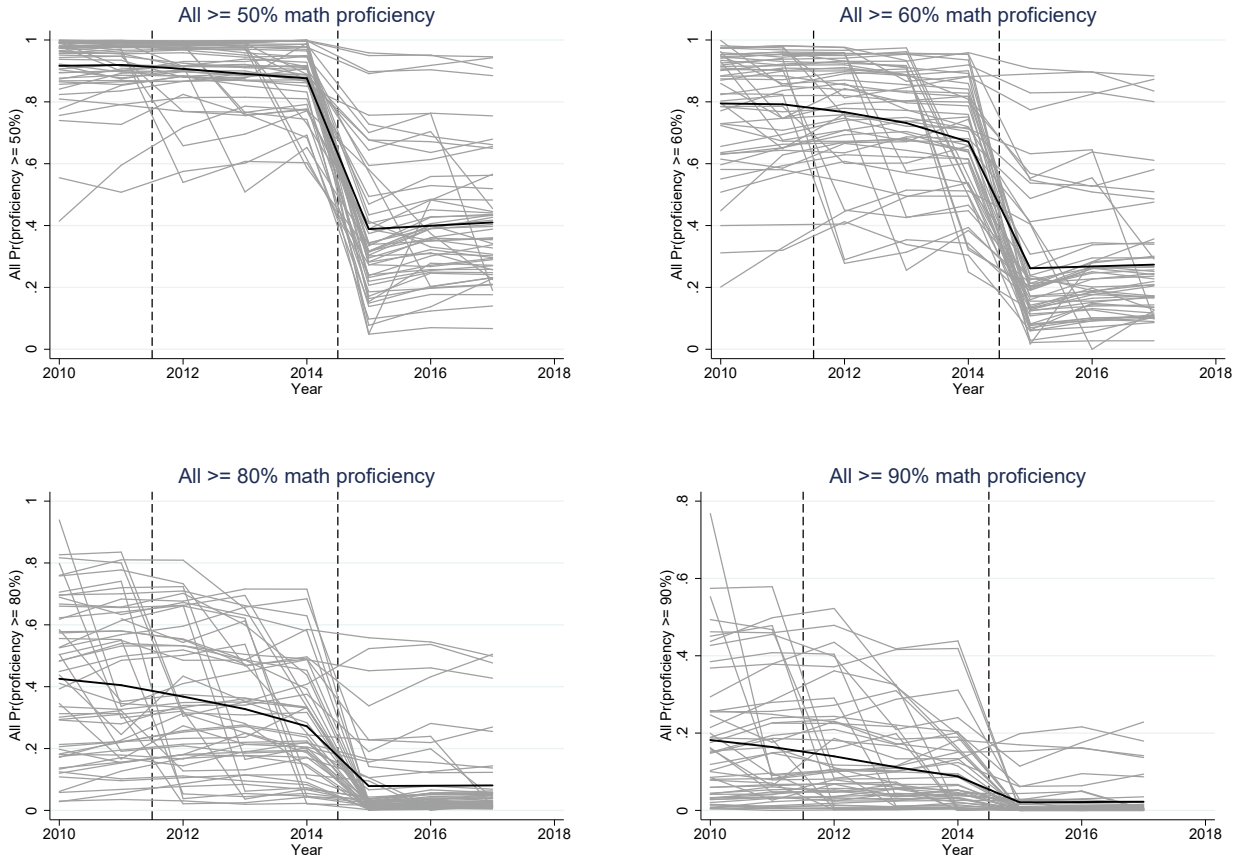
1. 2010-2013: Under the original No Child Left Behind (NCLB) legislation, schools were required to have all students scoring "proficient" by the 2014 school year (Institute of Education Sciences, 2009). As illustrated in Data Appendix Figure 1, there is very little variation in the proficiency measure over this period – in 29 states, most students were proficient in more than 95 percent of schools.

2. 2014: In the 2014 school year, 34 states received NCLB waivers, which provided exemptions from key elements of NCLB, including the proficiency requirements (US Department of Education, 2013)

3. 2015: States that received NCLB waivers are required to give "high-quality assessments" of "college- and career-ready standards." These examinations (e.g. PARCC and Smarter Balance) tended to have more stringent proficiency requirements than the earlier state examinations Education Week (2014). Accordingly, while there is within-state variation in the share of schools with proficiency marks upwards of 80 percent in the 2010-2013 period, there is little variation at these thresholds beginning in 2015 (probabilities of at least 80 percent students reaching proficiency falling close to 0, particularly for Hispanic and black students). In 2017, in 45 states, less than 5 percent of schools had achieved 90 percent proficiency (Data Appendix Figure 1).

These limitations are particularly challenging over the 2009 through 2015 period as public-use data only provide proficiency shares within wide performance bins (e.g.: at least 50 percent of students proficient), making it difficult to discern modest changes in performance.

Data Appendix Figure 1: Fraction of Students Achieving Math Proficiency, by State



Notes: Figures show the fraction of schools in each state with at least x% of students in each race/ethnic group achieving the state proficiency measure. Each line corresponds to one state; thick black line corresponds to the national average. Dashed vertical lines indicate 2012 (first pilot year of CEP) and 2015 (first year all states became eligible). All data from Department of Education EdFacts. Proficiency measures are available for schools with at least 6 (x = 50%), 16 (x = 60, 80%), or 31 (90%) students taking the examination.

Additional control variables

The SEDA data include a rich set of covariates for the geographic district (e.g.: including information from charter schools located in separate administrative districts, but the same geographic district as public schools). As with the performance data, (Reardon et al., 2018) provide a comprehensive description. From the SEDA data, I include control variables for the share of students in a district black, Hispanic, special education, or English Language Learners, as well as the fraction of students attending a charter school and student-teacher ratios derived from the CCD. Summary statistics for baseline economic variables, including median household income, the share of female-headed households, and the Gini coefficient, derived from the 2006-2010 pooled American Community Survey.

In addition to these covariates, I merge data to each school district from several outside sources. The Census Bureau Small Area Income and Poverty Estimates (SAIPE) program provides child poverty rates; the Bureau of Labor Statistics Local Area Unemployment

Statistics (LAUS) provides annual county unemployment rates. Data on district finances by level of government and type of revenue is provided by the LEA School District Finance Survey (F-33). Baseline rates of SNAP receipt are provided at the county level through USDA, and county-level per-capita income transfers are available through the Bureau of Economic Analysis Regional Economic Accounts (REIS) data. Finally, I explore heterogeneity by the local cost of living using the Bureau of Economic Analysis Regional Purchasing Parity (RPP) index.

I also augment the district CCD tabulations with several measures of *school* characteristics from the Common Core of Data. In particular, I estimate the fraction of black and Hispanic students attending CEP schools, and use these school-level counts to compute segregation measures for each race/ethnic group in a district-grade. The school district CCD data are also used to estimate the fraction of students gaining access to free meals under CEP relative to the traditional program (school-level data is necessary for this calculation for districts with partial CEP participation), as well as the continuous treatment measure presented in Appendix Tables 5 and 6.

Treatment definition

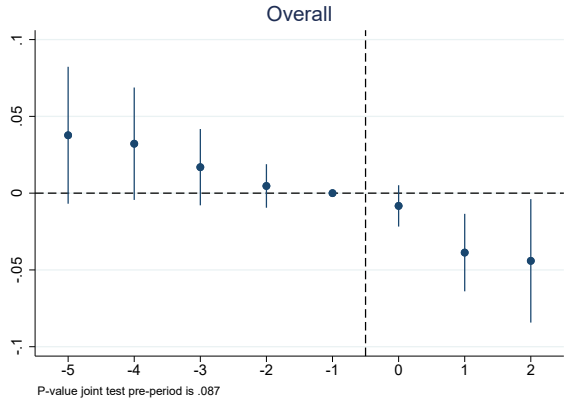
The analyses define CEP participation as the year in which a district-grade first participates in CEP and all subsequent years. Therefore, treatment depends on both the state in which a district is located and district-level decisions when to implement schoolwide free meals. In exploratory analyses, I have explored the feasibility of using the year each participating district became eligible for CEP (based on state) as the treatment variable. These analyses illustrate that using the timing of actual participation is better suited to evaluating the effect of CEP on student achievement than an eligibility-based treatment measure for two reasons.

First, calendar time and eligibility year are highly collinear: CEP was rolled out over a four-year period, with most schools (in 40 states) becoming eligible in 2015. Therefore, models using eligibility year as the treatment variable are unable to account for concurrent state policy changes that may affect both districts' (eventual) CEP participation and student achievement. In contrast, models that leverage actual implementation can account for these factors by augmenting the baseline specification with year-by-state fixed effects or state-specific linear trends. Appendix Figure 1 and Appendix Table 5 shows that math results are largely robust to these modifications. In contrast, reading results (Appendix Table 6) are more sensitive to the specification. Overall, it does not appear that CEP led to systematic changes in reading performance.

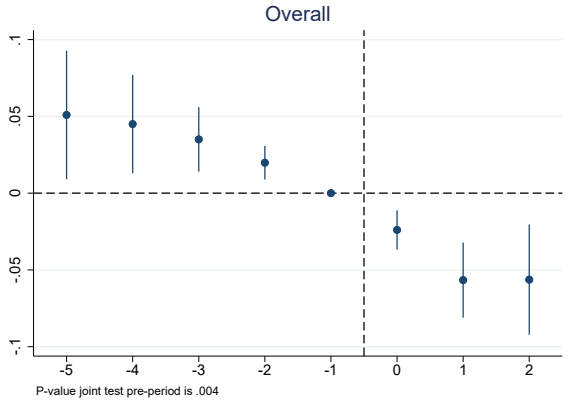
Second, it is possible that eligibility reflects selection at the state-level into treatment. If states were selected in part because of potential gains from CEP participation (or factors correlated with potential improvements), an eligibility-based treatment measure will not provide a biased measure of the causal effect of schoolwide free meals on student performance. Even with non-random selection of pilot states, however, the choice to participate in CEP is a school and district-level decision. The patterns shown in the eligibility-based event study plots in Data Appendix Figure 2 are consistent with negative selection of pilot states, although this is not conclusive evidence as it is difficult to disentangle selection from secular trends. In comparison, for black and Hispanic students, the timing of actual participation does not coincide with trends in student performance (Figure **??**).

Data Appendix Figure 2: Performance: Academic Performance Event Study: Eligbility-defined Treatment, Exposed Districts

(a) Math performance



(b) Reading performance



Notes: Figure presents results from event study framework in Equation 2 where event time is defined relative to the first year of CEP eligibility (2012 for districts in Michigan, Illinois, and Kentucky; 2013 for districts in the District of Columbia, New York, Ohio, and West Virginia; 2014 for districts in Florida, Georgia, Massachusetts, and Maryland; and 2015 for the remaining states). Outcome variable is overall math (panel (a)) or reading (panel (b)) performance subgroup results show similar patterns. All specifications include controls for student demographics, the fraction of charter schools in a district, child poverty and unemployment rates, measures of racial/ethnic segregation, year fixed effects, grade fixed effects, and district fixed effects. Bars denote 95 percent confidence intervals from robust standard errors clustered by district. Sample includes districts with a baseline FRP eligibility rate below 57.9 percent (the median among CEP-adopting districts). Notes below each panel present p-values from the joint test that pre-treatment coefficients equal to zero.

References

Education Week (2014). Big year looms for Common Core testing. https://www.edweek.org/ew/articles/2014/09/03/03assessment.h34.html.

Institute of Education Sciences (2009). Appendix A: State testing programs under NCLB. https://ies.ed.gov/ncee/pubs/2009013/appendix_a.asp.

Reardon, S. F., Ho, A. D., and Kalogrides, D. (2019). Validation methods for aggregate-level test scale linking: A case study mapping school district test score distributions to a common scale. *Stanford Center for Education Policy Analysis, Working Paper 16-09*.

Reardon, S. F., Shear, B. R., Castellano, K. E., and Ho, A. D. (2017). Using heteroskedastic ordered probit models to recover moments of continuous test score distributions from coarsened data. *Journal of Educational and Behavioral Statistics*, 42(1):3–45.

Reardon, S. M., Rahle, E. M., Shear, B. R., Kalogrides, D., DiSalvo, R., and Ho, A. D. (2018). Stanford education data archive: Technical documentation, version 2.1. https://stacks.stanford.edu/file/druid:db586ns4974/SEDA_documentation_v21.pdf.

US Department of Education (2013). States granted waivers from No Child Left Behind allowed to reapply for renewal for 2014 and 2015 school years. https://www.ed.gov/news/press-releases/states-granted-waivers-no-child-left-behind-allowed-reapply-renewal-2014-and-2015-sch