sociological science

Supplement to:

von Hippel, Paul T., and Caitlin Hamrock. 2019. "Do Test Score Gaps Grow Before, During, or Between the School Years? Measurement Artifacts and What We Can Know in Spite of Them." Sociological Science 6: 43-80.

Supplement

Multiple Imputation

All three datasets had missing test scores and covariates. Missing scores were especially noticeable in the GRD, where in the later grades some students with high fall scores were permitted to miss winter tests. Note that winter scores were missing at random (MAR) in the sense that the probability a winter score is missing depends on the fall score, which is observed. The MAR assumption justifies the multiple imputation procedure which we used to fill in missing scores (Allison, 2002).

We imputed each missing value 20 times, using a fully conditional specification in which each variable was imputed by regressing it on all the others (Raghunathan, Lepkowski, Van Hoewyk, & Solenberger, 2001). We used linear regression to impute continuous variables, and binomial, ordinal, and multinomial logit regression to impute categorical variables.

To account for correlations among tests given to the same student, before imputing we reshaped each dataset into wide format, so that there was one row per student and one column for each test occasion. This forced the imputation model to estimate correlations between different tests given to the same student (Allison, 2002). To account for correlations among students in the same school, we first imputed a school-level dataset containing school variables and school-level averages of child variables; then we joined the imputed school-level dataset to the child-level dataset and imputed the child-level variables.

Imputed test scores were used in graphs of mean test scores to improve estimation of trends. For our multilevel growth models, we used imputed covariates but deleted imputed test scores. This approach is known as *multiple imputation with deletion* (von Hippel, 2007). It produces slightly more efficient estimates of growth parameters and, more importantly, it reduces the sensitivity of the results to debatable choices made in the imputation model.

Multilevel growth model

We describe in detail the model that we fit to the GRD, which is the only dataset that tested children at least twice in every school year. Similar models were fit to the BSS and ECLS-K, with modification for school years with one test, or none.[1]

Our basic growth model was

$$Y_{sct} = \beta_0 + \alpha_0 Kinder_{sct} + \beta_1 Summer1_{sct} + \alpha_1 First_{sct} + \beta_2 Summer2_{sct} + \cdots + u_s + e_{sct}$$

where Y_{sct} was the reading or math score for child *c* in school *s* on measurement occasion *t*. All models were fit to the unstandardized original scores, as well as to scores that had been standardized and adjusted for reliability.

The exposure variables $Kinder_{sct}$, $Summer1_{sct}$, etc. were the number of months that the child had been exposed to each school year and summer (kindergarten, summer 1, etc.) as of measurement occasion *t*. These exposure variables were coded to reflect the fact that tests were not given on the first and last day of each school year; for example, at the time of the first test, the kindergarten exposure variable was approximately $Kinder_{sct}$=1.5 months, instead of 0.

The coefficients $\alpha_0, \alpha_1, \ldots$ were average monthly learning rates during each school year (kindergarten, first grade, etc.), and the coefficients $\beta_1, \beta_2, \ldots$ were average monthly learning rates

[1] In the GRD, tests were given in the fall and spring of every grade, so we could estimate the model as described with a separate growth rate for every school year and summer. In the ECLS-K, however, we could only estimate school and summer growth rates through the end of first grade, after which we could only estimate average growth rates over periods of 2-3 years. Similarly, in the BSS, we could only estimate school and summer growth rates from first through 5th grade; after 5th grade, we could only estimate 12-month growth rates, and before first grade we couldn't estimate anything because no tests were given.

rates during each summer (summer 1, summer 2, etc.). β_0 was the average score at the start of kindergarten. A school random effect u_s accounted for correlations among children from the same school, and the residual e_{sct} had an autocorrelated structure that accounted for the correlations among tests given to the same child.[2]

To estimate gaps in scores and score growth between advantaged and disadvantaged children, we added to the model a dummy variable X_{sc} which was 1 for advantaged children or schools and 0 for disadvantaged children or schools. In different model runs, X_{sc} was defined to indicate different measures of advantage—e.g., being white (vs. black), being nonpoor (vs. poor), or being in a low-poverty school (vs. a high-poverty school).

$$Y_{sct} = \beta_0 + \alpha_0 Kinder_{sct} + \beta_1 Summer1_{sct} + \alpha_1 First_{sct} + \beta_2 Summer2_{sct} + \cdots + \gamma_0 X_{sc} + \gamma_1 Kinder_{sct} X_{sc} + \delta_1 Summer1_{sct} X_{sc} + \gamma_2 First_{sct} X_{sc} + \delta_2 Summer2_{sct} X_{sc} + \cdots + u_s + e_{sct}$$

The coefficient γ_0 of X_{sc} represented the average score gap between advantaged and disadvantaged children at the start of kindergarten. We let X_{sc} interact with variables representing exposure to kindergarten, summer 1, etc. so that the coefficients $\gamma_1, \gamma_2, \ldots$ and $\delta_1, \delta_2, \ldots$ of the interactions represent monthly growth (or shrinkage) of the gap during each school year and summer.

Linear combinations of parameters

Our research questions were addressed not by the growth parameters themselves, but by linear combinations of the parameters.

We used linear combinations to estimate total growth in the gaps across all school years and summers from the beginning of first grade to the end of eighth—a period covered by all three datasets. On average, the school years were 9.37 months, and the summers were 2.63 months, so the total growth in the score gap was $\gamma_{1-8} = 9.37 \sum_{i=1}^{8} \gamma_i$ over the school years from first through eighth and $\delta_{2-8} = 2.63 \sum_{i=2}^{8} \delta_i$ over the summers between. Then the total growth in the gap from the start of first grade to the end of eighth grade was $\gamma_{1-8} + \delta_{2-8}$.

We also used linear combinations to compare the average monthly rate of gap growth during the school years and during summer vacations. In the GRD, the average monthly growth in the gap was $\bar{\gamma} = 1/9 \sum_{i=1}^{9} \gamma_i$ across the nine school years and $\bar{\delta} = 1/8 \sum_{i=1}^{8} \delta_i$ across the eight summers. So $\bar{\delta} - \bar{\gamma}$ was the average monthly difference between summer gap growth and school year gap growth. Wed defined similar contrasts for shorter periods such as grades K-1 in the ECLS-K or grades 1-6 in the BSS.

Supplemental results

Table A1 gives the mean and SD of each score on each test occasion from the fall of kindergarten (first grade in the BSS) through the spring of eighth grade.

Table A1. Means and SDs of test scales

[2] We modeled the serial correlation in different ways, first using a heteroscedastic AR1 model, and then using a spatial power model in which the residual correlation decreased as the time between tests increased (Littell, Milliken, Stroup, Wolfinger, & Schabenberger, 2006). The results were materially unchanged. An alternative way to account for correlations among tests on the same child is to include a child random intercept and possibly random slopes as well. But child random effects greatly increase runtime and memory use, to the point that SAS could not run models with child random effects on the GRD. The model for the GRD includes a fixed effect for each cohort, but these cohort fixed effects were small and made little difference to other estimates.

		BSS (CAT Thurstone scale)				ECLS-K (IRT ability θ)				ECLS-K (number right)				GRD (ability 10θ)			
		Reading		Math		Reading		Math		Reading		Math		Reading		Math	
Grade	Season	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
K	Fall					-1.31	.52	-1.16	.47	35	10	26	9	141	12	142	13
	Winter													149	13	149	14
	Spring					-.74	.51	-.68	.46	46	14	36	12	157	13	158	15
1	Fall	279	41	292	32	-.52	.51	-.45	.46	52	18	43	14	159	14	161	15
	Winter													168	15	171	15
	Spring	339	46	340	37	.10	.46	.04	.42	77	24	61	18	176	15	179	15
2	Fall	340	53	338	38									174	17	177	14
	Winter													182	16	185	14
	Spring	384	52	379	41									188	16	192	15
3	Fall	384	58	378	44									188	16	190	14
	Winter													194	16	198	14
	Spring	416	61	412	51	.78	.31	.71	.39	126	28	98	25	198	15	204	14
4	Fall	421	67	413	52									198	16	202	14
	Winter													203	15	208	14
	Spring	455	73	446	57									206	15	213	15
5	Fall	459	72	451	57									205	16	211	15
	Winter													209	15	217	16
	Spring	484	77	477	64	1.04	.30	1.09	.41	149	26	122	25	212	15	222	17
6	Fall	484	72	479	66									210	16	217	16
	Winter													213	15	222	16
	Spring	504	80	499	70									215	15	225	17
7	Fall													214	16	224	17
	Winter													217	16	227	18
	Spring	520	83	519	75									219	15	230	18
8	Fall													218	16	229	18
	Winter													220	16	233	19
	Spring	546	92	538	82	1.30	.39	1.42	.44	169	28	140	22	222	15	235	19

Table A2 estimates the scores' *reliability*. Most of the reliability estimates come from technical documentation published by the test developers (CTB/McGraw-Hill, 1979; Najarian et al., 2009; Northwest Evaluation Association, 2010).

For each scale, Table A2 presents two estimates of reliability. The first is an internal consistency estimate,[3] obtained by comparing a child's answers to different questions during the same administration of the same test form. The second is a test-retest estimate, obtained by correlating the scores on alternate forms of the same test taken a few days or weeks apart. Test-retest estimates are typically lower than internal consistency estimates of reliability because test-reliability estimates account for day-to-day as well as item-to-item variations in a child's performance (Shavelson & Webb, 1991). For some tests, the difference between test-retest and internal-consistency estimates was large; for other tests, it was small. For example, in the CAT test used by the BSS, the fall first grade reading scores were .68 reliable according to the internal consistency estimate but only .50 reliable according to the test-retest estimate. By contrast, the fall first grade math scores were .80-.81 reliable according to either estimate.

Strictly speaking, only the CAT offered a proper estimate of test-retest reliability, obtained by correlating scores from alternate test forms[4] (Forms C and D) taken two to three weeks apart (CTB/McGraw-Hill, 1979). For the GRD and ECLS-K, we did not have a proper estimate of test-reliability, but in the ECLS-K we did have correlations between fall and spring tests taken approximately six months apart, and in the GRD we had correlations between fall, winter, and spring tests taken approximately three to four months apart. These correlations (displayed in Table A2) were typically about .8-.9, and this figure may be viewed as a lower bound on reliability. It is a lower bound because six months is long enough for the true distribution of ability to change; that is, even if the tests were perfectly reliable, the correlation between fall and spring scores would be less than 1.0.

We adjusted for reliability using the internal-consistency reliability estimates from Table A2. The test-retest estimates led to very similar results.

[3] There are several ways to estimate reliability from internal consistency. Documentation for the CAT Form C used the KR-20 formula (CTB/McGraw-Hill, 1979; Kuder & Richardson, 1937), while the ECLS-K and GRD documentation used IRT formulas (Najarian, Pollack, & Sorongon, 2009; Northwest Evaluation Association, 2010).

[4] Alternate test forms were only used in grades 3-8; in grades 1-2 there were no alternate forms.

Table A2. Reliabilities of test scales

		BSS (CAT Thurstone scale)				ECLS-K (IRT ability scale θ)				GRD (IRT ability scale 10θ)			
		Reading		Math		Reading		Math		Reading		Math	
Grade	Season	Internal consistency	Test-retest	Internal consistency	Test-retest	Internal consistency	Corr. with next test	Internal consistency	Corr. with next test	Internal consistency	Corr. with next test	Internal consistency	Corr. with next test
K	Fall					.92	.80	.91	.83		.81		.82
	Winter									.90	.84	.91	.83
	Spring					.95	.89	.93	.90		.86		.85
1	Fall	.68	.50	.83	.80	.96	.83	.94	.82		.86		.85
	Winter									.93	.88	.94	.88
	Spring	.84		.87		.96	.76	.94	.73		.83		.82
2	Fall	.89	.73	.87	.80						.87		.87
	Winter									.96	.88	.96	.88
	Spring	.91		.90							.84		.84
3	Fall	.91	.78	.92	.83						.85		.85
	Winter									.95	.86	.95	.87
	Spring	.91		.93		.94	.86	.95	.85		.85		.86
4	Fall	.91	.80	.91	.83						.85		.88
	Winter									.94	.86	.95	.89
	Spring	.93		.93							.85		.88
5	Fall	.92	.81	.92	.83						.86		.89
	Winter									.94	.86	.96	.90
	Spring	.93		.94		.93	.79	.95	.78		.84		.88
6	Fall	.91	.75	.89	.77						.85		.89
	Winter									.94	.85	.96	.90
	Spring	.92		.91							.84		.90
7	Fall	.91	.78	.91	.76						.84		.90
	Winter									.94	.85	.97	.91
	Spring	.92		.92	.77						.83		.91
8	Fall	.88	.77	.91	.78						.84		.91
	Winter										.84	.97	.92
	Spring	.90		.92	.79	.87	.79	.92	.78				

Figures A1-A5 summarize gaps in math scores. These figures complement the reading gaps described by Figures 3-7 in the main text.

Tables A3-A4 estimate reading and math gaps at the start of first grade and the end of eighth grade, and total growth in those gaps over the period between. These are the estimates summarized in Table 3 of the main text.

Tables A5-A6 contrast the monthly growth (or shrinkage) of reading and math gaps during school years and summers over various periods: grades K-1, grades 1-6, and grades K-8. All these estimates were calculated using linear contrasts from our multilevel growth models. These are the estimates summarized in Table 4 of the main text.

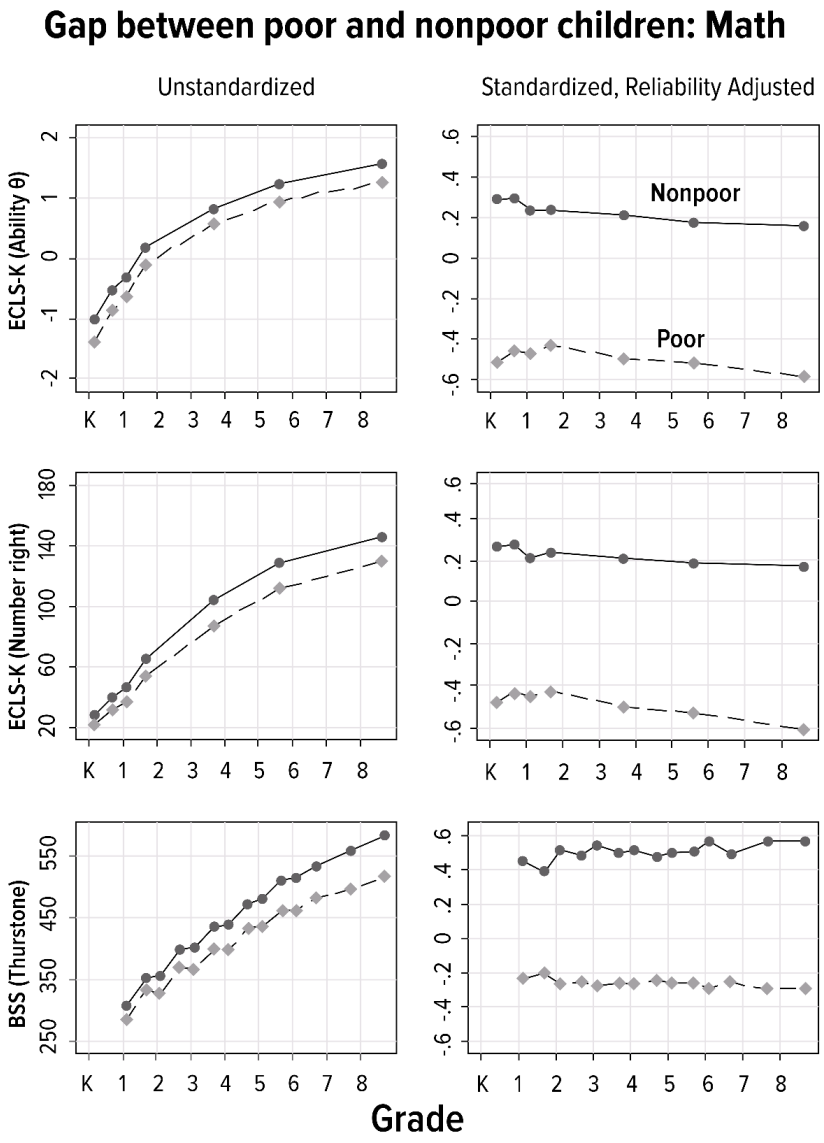

Figure A1. Mean math scores for poor and nonpoor children.

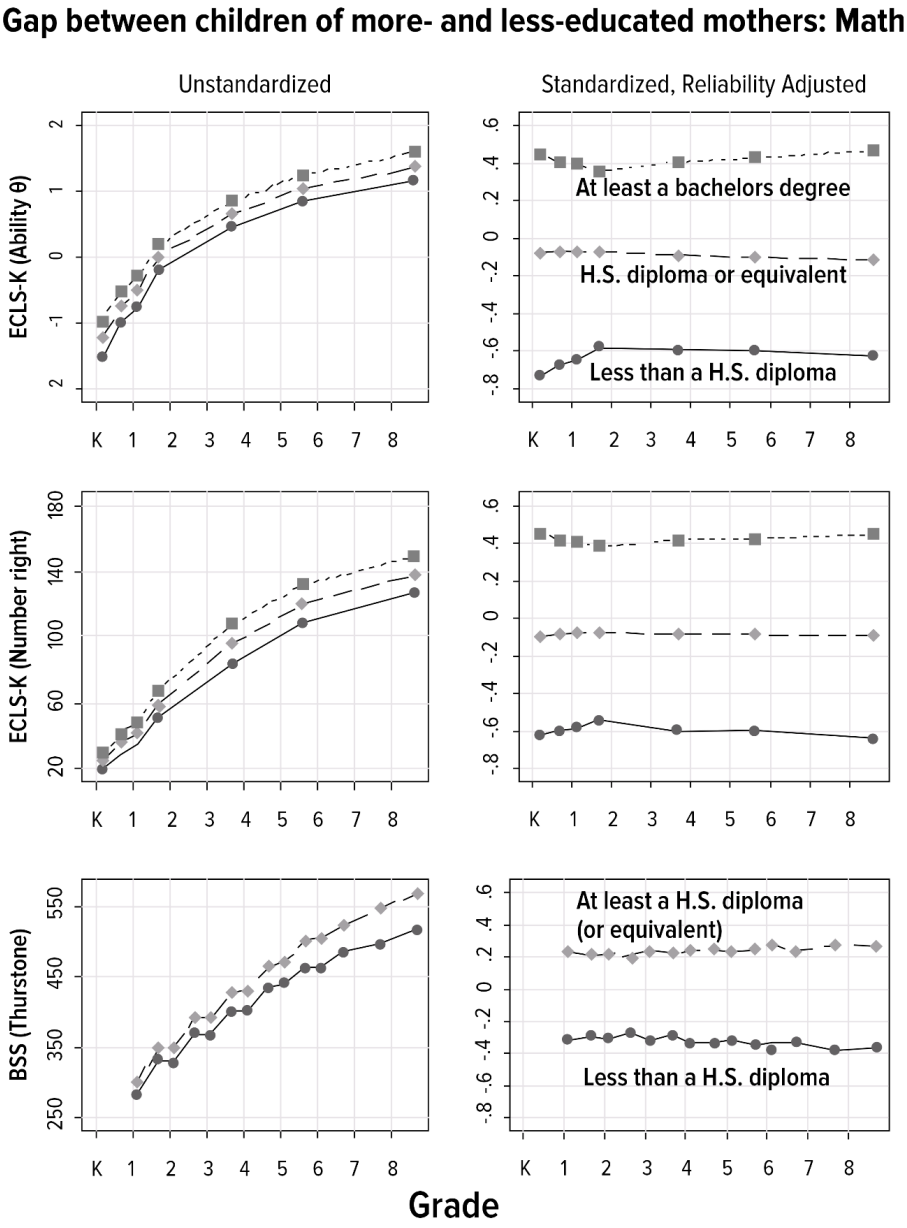

Figure A2. Mean math scores of children with more vs. less educated mothers.

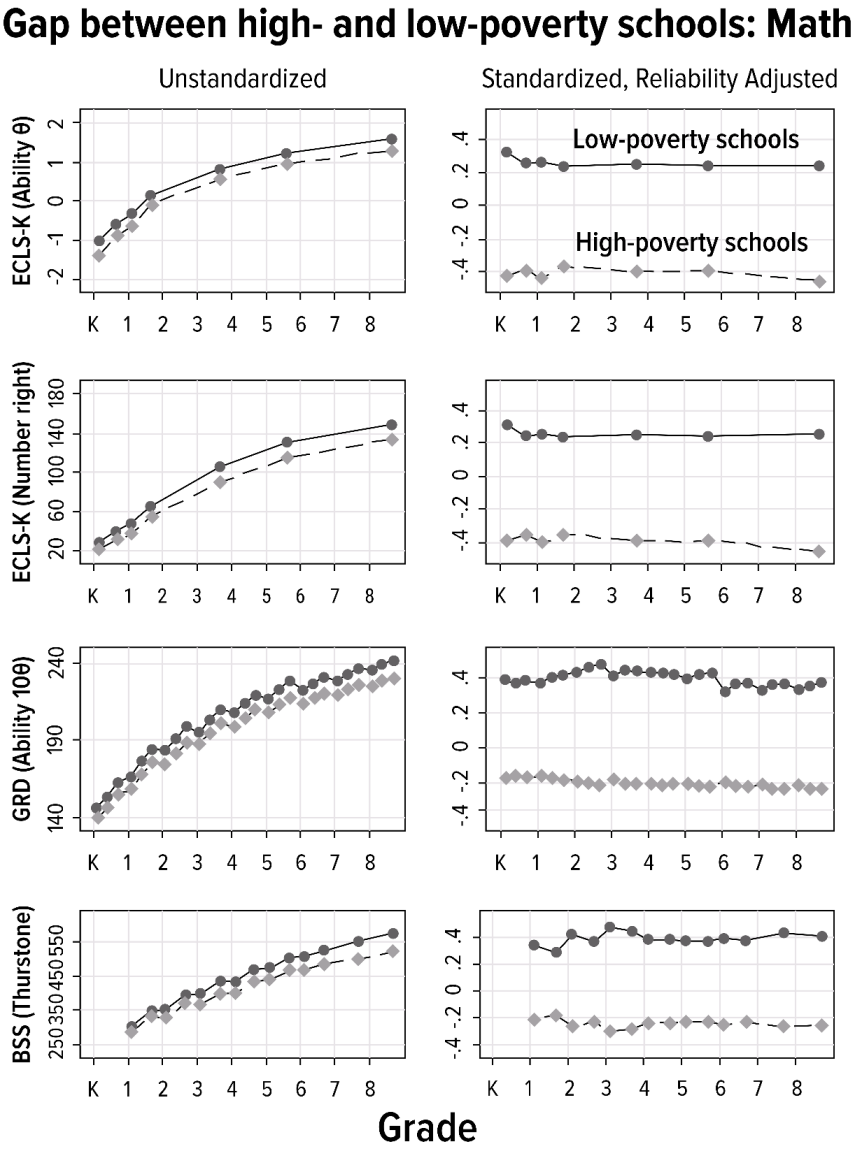

Figure A3. Mean math scores of children from low- and high-poverty schools.

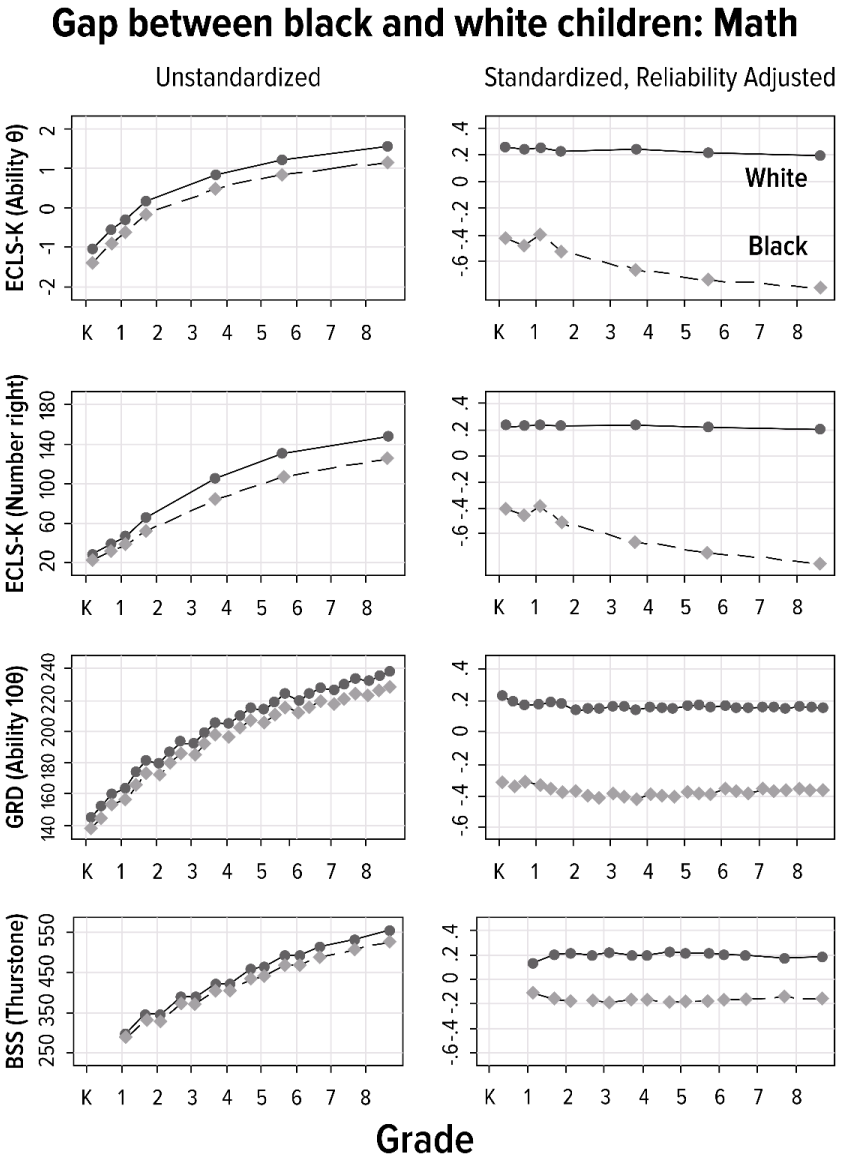

Figure A4. Mean math scores of white and black children.

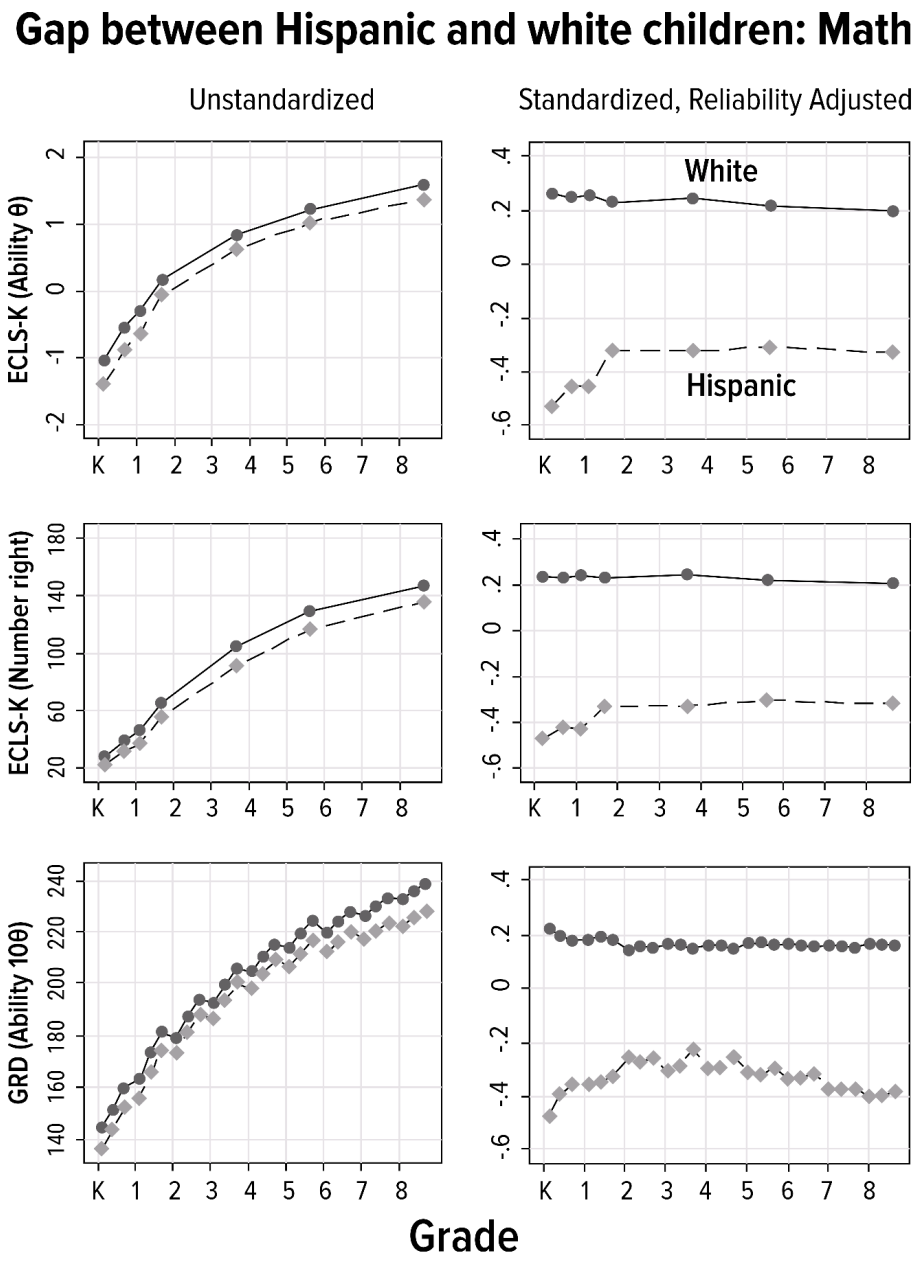

Figure A5. Mean math scores of white and Hispanic children.

Table A3. Reading gaps: Growth from first through eighth grade

a. Reading gap between nonpoor children and poor children

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.255***	0.187***	-0.067***	-26%	0.466***	0.492***	0.026	6%
	(0.010)	(0.008)	(0.008)		(0.022)	(0.021)	(0.015)	
ECLS-K (# right)	4.442***	14.396***	9.955***	224%	0.355***	0.525***	0.170***	48%
	(0.489)	(0.521)	(0.370)		(0.021)	(0.019)	(0.015)	
GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
BSS (Thurstone)	8.253*	66.988***	58.735***	712%	0.394**	0.662***	0.268†	68%
	(3.785)	(8.010)	(8.418)		(0.126)	(0.101)	(0.153)	

b. Reading gap between low-poverty and high-poverty schools

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
BSS (Thurstone)	16.330*	56.104***	39.774***	244%	0.586***	0.660***	0.074	13%
	(6.666)	(9.603)	(8.111)		(0.157)	(0.143)	(0.145)	
ECLS-K (# right)	7.122***	16.700***	9.578***	134%	0.451***	0.646***	0.195***	43%
	(0.677)	(0.592)	(0.519)		(0.032)	(0.029)	(0.021)	
ECLS-K (ability θ)	0.295***	0.256***	-0.039***	-13%	0.560***	0.630***	0.070**	13%
	(0.014)	(0.013)	(0.009)		(0.034)	(0.029)	(0.024)	
GRD (ability 10θ)	6.044***	7.930***	1.886***	31%	0.460***	0.506***	0.046	10%
	(0.389)	(0.448)	(0.418)		(0.026)	(0.031)	(0.028)	

c. Reading gap between children whose mothers did or did not have a high school diploma

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.291***	0.197***	-0.094***	-32%	0.515***	0.512***	-0.003	-1%
	(0.011)	(0.010)	(0.009)		(0.031)	(0.022)	(0.029)	
ECLS-K (# right)	3.621***	14.966***	11.345***	313%	0.348***	0.546***	0.198***	57%
	(0.793)	(0.633)	(0.640)		(0.030)	(0.022)	(0.026)	
GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
BSS (Thurstone)	8.229*	48.363***	40.135***	488%	0.401**	0.489***	0.089	22%
	(3.401)	(7.341)	(7.776)		(0.121)	(0.090)	(0.145)	

d. Reading gap between children whose mothers did or did not have a bachelor's degree

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.187***	0.138***	-0.049***	-26%	0.310***	0.351***	0.041*	13%
	(0.008)	(0.008)	(0.005)		(0.023)	(0.018)	(0.019)	
ECLS-K (# right)	2.844***	10.156***	7.312***	257%	0.275***	0.373***	0.099***	36%
	(0.681)	(0.633)	(0.445)		(0.022)	(0.018)	(0.018)	
GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
BSS (Thurstone)	NA	NA	NA	NA	NA	NA	NA	NA

e. Reading gap between white and black children

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.192***	0.234***	0.042***	22%	0.319***	0.577***	0.258***	81%
	(0.013)	(0.011)	(0.009)		(0.031)	(0.025)	(0.025)	
ECLS-K (# right)	2.537**	16.134***	13.597***	536%	0.226***	0.617***	0.391***	173%
	(0.796)	(0.645)	(0.516)		(0.029)	(0.026)	(0.018)	
GRD (ability 10θ)	3.948***	6.758***	2.810***	71%	0.326***	0.441***	0.115***	35%
	(0.277)	(0.369)	(0.456)		(0.019)	(0.025)	(0.031)	
BSS (Thurstone)	-3.719	16.942*	20.661*	-556%	-0.188	0.117	0.306*	-162%
	(3.984)	(8.147)	(8.148)		(0.124)	(0.106)	(0.144)	

f. Reading gap between white and Hispanic children

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.193***	0.153***	-0.040***	-21%	0.342***	0.350***	0.008	2%
	(0.012)	(0.010)	(0.009)		(0.032)	(0.023)	(0.028)	
ECLS-K (# right)	1.768†	10.849***	9.081***	514%	0.234***	0.398***	0.164***	70%
	(0.952)	(0.858)	(0.588)		(0.031)	(0.025)	(0.025)	
GRD (ability 10θ)	6.582***	7.982***	1.400**	21%	0.547***	0.528***	-0.020	-4%
	(0.273)	(0.362)	(0.451)		(0.019)	(0.025)	(0.031)	
BSS (Thurstone)	NA	NA						

*p<.05, **p<.01, ***p<.001. Growth and % Growth indicate how much a gap grows between first and eighth grade, in score points and in percentage points. NA indicates that a variable is not available in a particular dataset; for example, individual meal subsidy status is not available in the GRD.

Table A4. Math gaps: Growth from first through eighth grade

a. Math gap between nonpoor children and poor children

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	.240***	.175***	-.066***	-27%	.515***	.436***	-.079***	-15%
	(.010)	(.008)	(.007)		(.023)	(.019)	(.016)	
ECLS-K (# right)	5.564***	1.184***	4.619***	83%	.473***	.451***	-.021	-4%
	(.458)	(.443)	(.311)		(.021)	(.018)	(.017)	
GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
BSS (Thurstone)	12.023***	56.436***	44.412***	369%	.534***	.663***	.129	24%
	(2.833)	(6.881)	(6.955)		(.097)	(.100)	(.127)	

b. Math gap between low-poverty and high-poverty schools

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
BSS (Thurstone)	14.889**	47.320***	32.432***	218%	0.582***	0.632***	0.050	9%
	(5.622)	(8.287)	(6.699)		(0.140)	(0.144)	(0.122)	
ECLS-K (# right)	7.923***	12.145***	4.223***	53%	0.593***	0.584***	-0.009	-2%
	(0.572)	(0.491)	(0.411)		(0.032)	(0.029)	(0.018)	
ECLS-K (ability θ)	0.296***	0.242***	-0.054***	-18%	0.629***	0.572***	-0.058**	-9%
	(0.014)	(0.012)	(0.009)		(0.034)	(0.028)	(0.022)	
GRD (ability 10θ)	5.918***	11.630***	5.712***	97%	0.397***	0.575***	0.178***	45%
	(0.445)	(0.496)	(0.412)		(0.029)	(0.033)	(0.027)	

c. Math gap between children whose mothers did or did not have a h.s. diploma (or equivalent)

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.245***	0.161***	-0.084***	-34%	0.523***	0.404***	-0.119***	-23%
	(0.012)	(0.009)	(0.011)		(0.030)	(0.021)	(0.027)	
ECLS-K (# right)	4.785***	9.690***	4.905***	103%	0.444***	0.425***	-0.019	-4%
	(0.629)	(0.468)	(0.547)		(0.027)	(0.022)	(0.022)	
GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
BSS (Thurstone)	8.552***	39.183***	30.632***	358%	0.431***	0.469***	0.038	9%
	(2.354)	(6.615)	(6.626)		(0.089)	(0.093)	(0.119)	

d. Math gap between children whose mothers did or did not have a bachelor's degree

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.182***	0.141***	-0.041***	-23%	0.374***	0.346***	-0.028	-8%
	(0.007)	(0.007)	(0.005)		(0.019)	(0.015)	(0.016)	
ECLS-K (# right)	4.180***	7.733***	3.552***	85%	0.396***	0.372***	-0.024*	-6%
	(0.466)	(0.426)	(0.346)		(0.017)	(0.018)	(0.011)	
GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
BSS (Thurstone)	NA	NA	NA	NA	NA	NA	NA	NA

e. Math gap between white and black children

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.267***	0.283***	0.016†	6%	0.552***	0.679***	0.127***	23%
	(0.014)	(0.011)	(0.010)		(0.031)	(0.024)	(0.027)	
ECLS-K (# right)	6.277***	14.894***	8.617***	137%	0.528***	0.714***	0.186***	35%
	(0.635)	(0.448)	(0.520)		(0.029)	(0.026)	(0.020)	
GRD (ability 10θ)	6.569***	9.672***	3.103***	47%	0.462***	0.490***	0.028	6%
	(0.272)	(0.362)	(0.447)		(0.018)	(0.024)	(0.030)	
BSS (Thurstone)	9.125**	24.798***	15.673*	172%	0.156	0.211*	0.056	36%
	(2.998)	(6.879)	(6.651)		(0.098)	(0.105)	(0.120)	

f. Math gap between white and Hispanic children

	Unstandardized				Standardized, reliability-adjusted			
Dataset (scale)	Start first	End eighth	Growth	% growth	Start first	End eighth	Growth	% growth
ECLS-K (ability θ)	0.241***	0.150***	-0.091***	-38%	0.533***	0.350***	-0.183***	-34%
	(0.011)	(0.010)	(0.008)		(0.028)	(0.021)	(0.024)	
ECLS-K (# right)	5.587***	8.230***	2.643***	47%	0.507***	0.381***	-0.126***	-25%
	(0.571)	(0.420)	(0.476)		(0.026)	(0.023)	(0.018)	
GRD (ability 10θ)	7.196***	8.753***	1.558***	22%	0.531***	0.433***	-0.098***	-18%
	(0.262)	(0.349)	(0.434)		(0.017)	(0.023)	(0.029)	
BSS (Thurstone)	NA	NA	NA	NA	NA	NA	NA	NA

*p<.05, **p<.01, ***p<.001. Growth and % Growth indicate how much a gap grows between first and eighth grade, in score points and in percentage points. NA indicates that a variable is not available in a particular dataset; for example, individual meal subsidy status is not available in the GRD.

Table A5. Reading gap growth, per month, during school and summer

a. Monthly reading gap growth between nonpoor children and poor children

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.007***	0.012***	0.019***	Summer	-0.009***	0.015**	0.024***	Summer
		(0.001)	(0.002)	(0.003)		(0.001)	(0.004)	(0.005)	
	ECLS-K (# right)	0.530***	0.229*	-0.301**	School	0.000	-0.002	-0.002	NS
		(0.024)	(0.097)	(0.111)		(0.001)	(0.004)	(0.005)	
	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
1-6	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
	BSS (Thurstone)	0.195	2.111**	1.916*	Summer	-0.005	0.033*	0.038*	Summer
		(0.195)	(0.700)	(0.850)		(0.004)	(0.014)	(0.017)	
K-8	GRD (ability 10θ)	NA	NA						

b. Monthly reading gap growth between low-poverty and high-poverty schools

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.007***	0.017***	0.024***	Summer	-0.010***	0.026***	0.036***	Summer
		(0.000)	(0.002)	(0.003)		(0.002)	(0.007)	(0.008)	
	ECLS-K (# right)	0.470***	0.300*	-0.169	NS	-0.002	0.007	0.009	NS
		(0.032)	(0.152)	(0.177)		(0.001)	(0.006)	(0.007)	
	GRD (ability 10θ)	0.168***	-0.009	-0.177*	School	0.008***	-0.019***	-0.027***	School
		(0.021)	(0.079)	(0.090)		(0.001)	(0.005)	(0.006)	
1-6	GRD (ability 10θ)	-0.008	0.149***	0.157***	Summer	0.002***	-0.007***	-0.009***	School
		(0.008)	(0.029)	(0.035)		(0.001)	(0.002)	(0.002)	
	BSS (Thurstone)	-0.366†	3.428***	3.794***	Summer	-0.014***	0.057***	0.071***	Summer
		(0.189)	(0.668)	(0.813)		(0.004)	(0.014)	(0.017)	
K-8	GRD (ability 10θ)	0.040***	0.023	-0.017	NS	0.005***	-0.014***	-0.019***	School
		(0.008)	(0.024)	(0.029)		(0.001)	(0.002)	(0.002)	

c. Monthly reading gap growth between children whose mothers did or did not have a h.s. diploma (or equivalent)

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.007***	0.020***	0.026***	Summer	-0.007***	0.029***	0.036***	Summer
		(0.000)	(0.002)	(0.003)		(0.002)	(0.009)	(0.010)	
	ECLS-K (# right)	0.613***	0.343	-0.270	NS	0.003	0.005	0.002	NS
		(0.040)	(0.198)	(0.228)		(0.002)	(0.008)	(0.009)	
	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
1-6	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
	BSS (Thurstone)	0.246	0.981	0.734	NS	-0.003	0.013	0.016	NS
		(0.196)	(0.702)	(0.861)		(0.004)	(0.015)	(0.018)	
K-8	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA

d. Monthly reading gap growth between children whose mothers did or did not have a bachelor's degree

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows faster
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	in
K-1	ECLS-K (ability θ)	-0.006***	0.008***	0.015***	Summer	-0.008***	0.007	0.015*	Summer
		(0.000)	(0.001)	(0.002)		(0.001)	(0.006)	(0.007)	
	ECLS-K (# right)	0.444***	0.132	-0.313	NS	-0.001	-0.004	-0.003	NS
		(0.029)	(0.139)	(0.160)		(0.001)	(0.005)	(0.006)	
	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
1-6	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
	BSS (Thurstone)	NA	NA	NA	NA	NA	NA	NA	NA
K-8	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA

e. Monthly reading gap growth between white and black children

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	0.001**	0.001	0.000	NS	0.008***	-0.011	-0.019*	School
		(0.000)	(0.003)	(0.003)		(0.001)	(0.007)	(0.008)	
	ECLS-K (# right)	0.578***	-0.111	-0.689***	School	0.012***	-0.022***	-0.034***	School
		(0.031)	(0.149)	(0.171)		(0.001)	(0.004)	(0.005)	
	GRD (ability 10θ)	0.193***	-0.389***	-0.581***	School	0.005*	-0.026***	-0.030***	School
		(0.028)	(0.105)	(0.119)		(0.002)	(0.007)	(0.008)	
1-6	GRD (10θ)	0.070***	-0.079*	-0.149***	School	0.005***	-0.014***	-0.019***	School
		(0.010)	(0.036)	(0.044)		(0.001)	(0.002)	(0.003)	
	BSS (Thurstone)	0.044	1.307*	1.262	NS	0.001	0.019	0.017	NS
		(0.182)	(0.643)	(0.783)		(0.004)	(0.013)	(0.016)	
K-8	GRD (ability 10θ)	0.077***	-0.134***	-0.211***	School	0.005***	-0.016***	-0.020***	School
		(0.010)	(0.030)	(0.038)		(0.001)	(0.002)	(0.003)	

f. Monthly reading gap growth between white and Hispanic children

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.008***	0.010***	0.018***	Summer	-0.014***	0.025**	0.039***	Summer
		(0.001)	(0.002)	(0.003)		(0.002)	(0.009)	(0.010)	
	ECLS-K (# right)	0.447***	-0.001	-0.448*	School	-0.003	0.003	0.005	NS
		(0.038)	(0.189)	(0.218)		(0.002)	(0.007)	(0.009)	
	GRD (ability 10θ)	0.130***	-0.123	-0.254*	School	-0.005**	0.004	0.010	NS
		(0.028)	(0.100)	(0.115)		(0.002)	(0.007)	(0.008)	
1-6	GRD (ability 10θ)	0.014	-0.047	-0.060	NS	-0.002*	-0.002	0.000	NS
		(0.010)	(0.036)	(0.044)		(0.001)	(0.002)	(0.003)	
	BSS (Thurstone)	NA	NA	NA	NA	NA	NA	NA	NA
K-8	GRD (ability 10θ)	0.041***	-0.051	-0.092*	School	-0.001*	0.001	0.003	NS
		(0.009)	(0.029)	(0.037)		(0.001)	(0.002)	(0.003)	

*p<.05, **p<.01, ***p<.001. NS means p>.05. NA means that the variable is not available in a particular dataset. For example, there are no Hispanics in the BSS and no data on income or maternal education in the GRD.

Table A6. Math gaps: Monthly gap growth rates during school and summer

a. Monthly math gap growth between nonpoor children and poor children

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.008***	0.009***	0.016***	Summer	-0.012***	0.014**	0.026***	Summer
		(0.000)	(0.002)	(0.003)		(0.001)	(0.005)	(0.006)	
	ECLS-K (# right)	0.273***	0.327***	0.054	NS	-0.008***	0.010*	0.018**	Summer
		(0.018)	(0.086)	(0.100)		(0.001)	(0.005)	(0.006)	
	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
1-6	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
	BSS (Thurstone)	0.188	1.420**	1.232*	Summer	-0.010**	0.047***	0.057***	Summer
		(0.139)	(0.475)	(0.571)		(0.003)	(0.012)	(0.014)	
K-8	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA

b. Monthly math gap growth between low-poverty and high-poverty schools

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.007***	0.009***	0.016***	Summer	-0.012***	0.014*	0.026***	Summer
		(0.001)	(0.003)	(0.003)		(0.001)	(0.007)	(0.008)	
	ECLS-K (# right)	0.239***	0.338**	0.099	NS	-0.009***	0.011*	0.020***	Summer
		(0.026)	(0.123)	(0.142)		(0.001)	(0.005)	(0.006)	
	GRD (ability 10θ)	0.088***	0.241**	0.153	Summer	0.006***	-0.005	-0.011	School
		(0.020)	(0.076)	(0.086)		(0.001)	(0.005)	(0.006)	
1-6	GRD (ability 10θ)	0.113***	-0.213***	-0.326***	School	0.005***	-0.011***	-0.016***	School
		(0.008)	(0.028)	(0.034)		(0.001)	(0.002)	(0.002)	
	BSS (Thurstone)	0.223†	0.728	0.505	NS	-0.008*	0.034**	0.042**	Summer
		(0.133)	(0.453)	(0.545)		(0.003)	(0.012)	(0.014)	
K-8	GRD (ability 10θ)	0.130***	-0.155***	-0.285***	School	0.005***	-0.011***	-0.017***	School
		(0.007)	(0.023)	(0.028)		(0.000)	(0.002)	(0.002)	

c. Monthly math gap growth between children whose mothers only completed high school and children whose mothers dropped out

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.008***	0.008**	0.016***	Summer	-0.013***	0.014†	0.027**	Summer
		(0.001)	(0.003)	(0.004)		(0.002)	(0.008)	(0.009)	
	ECLS-K (# right)	0.288***	0.314†	0.026	NS	-0.007***	0.006	0.013†	NS
		(0.035)	(0.171)	(0.197)		(0.001)	(0.006)	(0.007)	
	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
1-6	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
	BSS (Thurstone)	0.268*	0.419	0.150	NS	-0.005	0.019	0.023	NS
		(0.134)	(0.474)	(0.571)		(0.003)	(0.012)	(0.014)	
K-8	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA

d. Monthly math gap growth between children whose mothers went beyond high school and children whose mothers completed high school only

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.005***	0.006***	0.011***	Summer	-0.007***	0.007	0.015**	Summer
		(0.000)	(0.001)	(0.002)		(0.001)	(0.005)	(0.006)	
	ECLS-K (# right)	0.213***	0.311**	0.098	NS	-0.008***	0.013***	0.021***	Summer
		(0.021)	(0.100)	(0.115)		(0.001)	(0.002)	(0.003)	
	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
1-6	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA
	BSS (Thurstone)	NA	NA	NA	NA	NA	NA	NA	NA
K-8	GRD (ability 10θ)	NA	NA	NA	NA	NA	NA	NA	NA

e. Monthly math gap growth between white and black children

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows faster in
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	
K-1	ECLS-K (ability θ)	0.000	0.000	-0.001	NS	0.006***	-0.012	-0.018*	School
		(0.001)	(0.003)	(0.003)		(0.002)	(0.008)	(0.009)	
	ECLS-K (# right)	0.475***	0.121	-0.353*	School	0.008***	-0.010*	-0.018**	School
		(0.032)	(0.154)	(0.177)		(0.001)	(0.005)	(0.006)	
	GRD (ability 10θ)	0.114***	0.055	-0.060	NS	0.001	0.001	0.000	NS
		(0.027)	(0.100)	(0.113)		(0.002)	(0.007)	(0.007)	
1-6	GRD (ability 10θ)	0.127***	-0.358***	-0.485***	School	0.004***	-0.013***	-0.017***	School
		(0.010)	(0.035)	(0.043)		(0.001)	(0.002)	(0.003)	
	BSS (Thurstone)	0.327**	-0.211	-0.537	NS	0.001	0.002	0.000	NS
		(0.126)	(0.422)	(0.509)		(0.003)	(0.011)	(0.014)	
K-8	GRD (ability 10θ)	0.108***	-0.235***	-0.343***	School	0.001	-0.007***	-0.008**	School
		(0.009)	(0.029)	(0.036)		(0.001)	(0.002)	(0.002)	

f. Monthly math gap growth between white and Hispanic children

		Unstandardized			Gap grows	Standardized, reliability-adjusted			Gap grows
Grades	Dataset (scale)	School	Summer	Difference	faster in	School	Summer	Difference	faster in
K-1	ECLS-K (ability θ)	-0.008***	0.002	0.011***	Summer	-0.017***	0.011	0.028***	Summer
		(0.000)	(0.002)	(0.003)		(0.001)	(0.007)	(0.008)	
	ECLS-K (# right)	0.241***	0.170	-0.071	NS	-0.011***	0.009*	0.020***	Summer
		(0.030)	(0.146)	(0.168)		(0.001)	(0.005)	(0.005)	
	GRD (ability 10θ)	0.009	0.262**	0.254*	Summer	-0.010***	0.027***	0.038***	Summer
		(0.025)	(0.093)	(0.106)		(0.002)	(0.006)	(0.007)	
1-6	GRD (ability 10θ)	0.075***	-0.312***	-0.388***	School	-0.003***	0.003	0.006*	Summer
		(0.009)	(0.035)	(0.042)		(0.001)	(0.002)	(0.003)	
	BSS (Thurstone)	NA	NA	NA	NA	NA	NA	NA	NA
K-8	GRD (ability 10θ)	0.074***	-0.193***	-0.267***	School	-0.005***	0.011***	0.016***	Summer
		(0.009)	(0.028)	(0.035)		(0.001)	(0.002)	(0.002)	

*p<.05, **p<.01, ***p<.001. NA means that the variable or grade range is not available in a particular dataset.