Online Appendix

for

School Effects on Socio-emotional Development, School-Based Arrests, and Educational Attainment

C. Kirabo Jackson, Shanette C. Porter

John Q. Easton, Alyssa Blanchard, and Sebastián Kiguel

Supplementary Material

Table S1: Testing Selection on Observables

	1 Math 8	2 ELA 8	3 Emotional Health 8	4 Academic Engagement 8	5 Grit 8	6 School Connectedness 8	7 Study Habits 8	8 Absences 8	9 GPA 8	10 Incidents 8	11 Suspensions 8
Social Value-Added	0.0261	0.0267	0.0102	0.00695	0.0187	0.0106	0.0144	-0.0660	0.00207	0.000305	0.0148
	(0.0400)	(0.0364)	(0.00801)	(0.00650)	(0.0129)	(0.0115)	(0.0106)	(0.123)	(0.0234)	(0.00372)	(0.0311)
Work Hard Value-Added	0.0281	0.0156	0.00513	0.00430	0.00958	0.0209	0.0120	0.156	0.0165	-0.00116	0.0118
	(0.0578)	(0.0574)	(0.0117)	(0.0127)	(0.0233)	(0.0176)	(0.0158)	(0.241)	(0.0407)	(0.00665)	(0.0568)
Test Scores Value-Added	-0.00438	0.00321	-0.00750	0.00634	0.00196	-0.00533	0.00430	-0.193	0.00748	-9.51e-05	-0.0314
	(0.0643)	(0.0610)	(0.0117)	(0.0127)	(0.0184)	(0.0159)	(0.0151)	(0.202)	(0.0403)	(0.00698)	(0.0601)
	12 Female	13 Special Ed	14 Free Lunch	15 Reduced Lunch	16 White	17 Black	18 Native	19 Asian/Pacific Islander	20 Latinx	21 Multiracial	22 Asian
Social Value-Added	0.00380	-0.00174	-0.0133	0.000191	0.00548	0.0225	-7.25e-05	-0.000168	-0.0284	0.000264	-0.000522
	(0.00473)	(0.00327)	(0.0115)	(0.00259)	(0.00775)	(0.0208)	(0.000158)	(0.00295)	(0.0199)	(0.000197)	(0.000823)
Work Hard Value-Added	-0.000642	-0.00165	-0.00750	-0.00191	0.0122	-0.00180	-5.67e-06	0.00864	-0.0198	0.000185	0.00117
	(0.0114)	(0.00708)	(0.0203)	(0.00514)	(0.0170)	(0.0448)	(0.000341)	(0.00553)	(0.0394)	(0.000348)	(0.00204)
Test Scores Value-Added	0.00211	0.00665	0.000850	0.00153	-0.00540	0.00134	0.000234	-0.00671	0.0126	-0.000198	-0.00143
	(0.0111)	(0.00583)	(0.0229)	(0.00415)	(0.0166)	(0.0402)	(0.000297)	(0.00580)	(0.0347)	(0.000364)	(0.00200)
	23 Math 7	24 ELA 7	25 Emotional Health 7	26 Academic Engagement 7	27 Grit 7	28 School Connectedness 7	29 Study Habits 7	30 Absences 7	31 GPA 7	32 Incidents 7	33 Suspensions 7
Social Value-Added	0.0242	0.0287	0.00178	0.00289	0.0122	0.00548	0.0106	-0.224	0.0127	0.00100	0.00919
	(0.0391)	(0.0370)	(0.00519)	(0.00393)	(0.00930)	(0.00572)	(0.00853)	(0.453)	(0.0274)	(0.00276)	(0.0290)
Work Hard Value-Added	0.0249	0.0120	0.0107	0.00104	0.00605	0.0135	0.0115	0.188	0.0159	0.00317	0.0190
	(0.0596)	(0.0599)	(0.00889)	(0.00829)	(0.0170)	(0.00949)	(0.0114)	(0.848)	(0.0473)	(0.00531)	(0.0561)
Test Scores Value-Added	-0.00578	0.00109	-0.0105	-0.00165	0.00260	-0.00639	-0.00491	-0.459	-0.000230	-0.00387	-0.0404
	(0.0659)	(0.0636)	(0.00864)	(0.00794)	(0.0139)	(0.00901)	(0.0110)	(0.795)	(0.0477)	(0.00561)	(0.0586)

Robust standard errors in parentheses adjusted for clustering at the school level.

Note: Results are based on regression of each covariate on out-of-sample Social, Work hard, and Test Scores Value Added, and year fixed effects. No controls are included in these models. Missing 8^{th} grade measures were imputed using 7^{th} grade measures and demographic characteristics. The sample includes all students in neighborhood, charter, and magnet schools, between 2011 and 2017. To summarize the result in this table, we also estimate effects on "predicted" measures and outcomes based on all of the covariates in this table. The results are in Table S2.

Table S2: Testing Selection on Observables: Effects on Predicted Outcomes and Measures

	(1) Predicted: High School Graduation	(2) Predicted: College Enrollment	(3) Predicted: 4-Year College Enrollment	(4) Predicted: 2-Year College Enrollment	(5) Predicted: Work Hard	(6) Predicted: Social	(7) Predicted: Test Scores
Workhard Value-Added	0.000654	0.00437	0.00677	-0.00122	0.000166	0.00305	0.0102
	(0.00956)	(0.0142)	(0.0155)	(0.00392)	(0.0116)	(0.0123)	(0.0545)
Social Value-Added	0.00325	0.00135	-0.000394	0.00183	0.00312	0.00161	0.00179
	(0.00926)	(0.0146)	(0.0159)	(0.00347)	(0.0103)	(0.0138)	(0.0635)
Test Scores Value-Added	0.00184	0.00530	0.00818	-0.00132	0.0144	0.0142	0.0515
	(0.00547)	(0.00869)	(0.00974)	(0.00191)	(0.00758)	(0.00925)	(0.0419)
Observations	55,560	55,560	55,560	55,560	160,148	160,148	160,148
P-value from F-statistic on all Value-Addeds	0.9295	0.8742	0.7776	0.8740	0.2341	0.3612	0.5638
R2 of predicted outcome on actual outcome	0.1751	0.2199	0.2374	0.0382	0.2002	0.1438	0.6089

Robust standard errors in parentheses adjusted for clustering at the school level.

Predicted outcomes are fitted values from a linear regression of said outcome on *all* observed covariates listed in Table S1. To avoid mechanical correlation, we use leave-year out predicted outcomes (i.e., predicted outcomes based on the relationship between the outcome and covariates in *other* years). The R-squared of predicted outcomes on actual outcome is reported in the bottom row. The reported point estimates are those on predicted outcomes on the value-addeds with no controls. The *p*-value associated with the joint significance of all three value-addeds is reported in the second to the lowest row. Missing 8^{th} grade measures were imputed using 7^{th} grade measures and demographic characteristics. for the longer-run outcomes, the sample includes first time 9^{th} grade students between 2011 and 2014. For the measures, the sample includes first time 9^{th} grade students between 2011 and 2017.

Table S3: Summary Statistics for Survey Completers and Non-Completers

	Full Sample		Completed Surveys in 9th Grade		Did Not Complete Surveys in 9th Grade	
Variables	mean	SD	mean	SD	mean	SD
9th grade						
Math in 9th Grade	0.0101	(0.991)	0.0630	(0.975)	-0.162	(1.023)
English in 9th Grade	0.00459	(0.995)	0.0605	(0.986)	-0.184	(1.001)
Emotional health in 9th Grade	-0.0529	(0.989)	-0.0473	(0.987)	-0.219	(1.043)
Academic engagement in 9th Grade	0.113	(1.061)	0.113	(1.060)	0.138	(1.111)
Grit in 9th Grade	-0.0122	(0.950)	-0.00633	(0.946)	-0.199	(1.035)
School connectness in 9th Grade	-0.0351	(0.956)	-0.0313	(0.953)	-0.138	(1.042)
Study habits in 9th Grade	-0.0178	(1.004)	-0.0132	(1.003)	-0.175	(1.045)
Absences in 9th Grade	14.998	(18.633)	12.9	(15.538)	21.49	(24.875)
GPA in 9th Grade	2.426	(1.007)	2.503	(0.974)	2.154	(1.071)
Days Suspended in 9th Grade	0.804	(3.293)	0.633	(2.772)	1.334	(4.501)
Incidents in 9th Grade	0.0769	(0.419)	0.0607	(0.356)	0.127	(0.568)
On Track in 9th Grade	0.848	(0.359)	0.872	(0.334)	0.759	(0.428)
Indices						
Test Scores in 9th Grade	-0.00746	(0.992)	0.0485	(0.977)	-0.191	(1.018)
Workhard in 9th Grade	0.182	(0.988)	0.190	(0.981)	-0.0221	(1.128)
Social in 9th Grade	0.00430	(0.998)	0.00975	(0.994)	-0.141	(1.095)
8th grade						
Math in 8th Grade	0.202	(0.938)	0.261	(0.930)	0.0202	(0.939)
English in 8th Grade	0.207	(0.937)	0.267	(0.916)	0.0197	(0.974)
Emotional health in 8th Grade	0.0704	(0.896)	0.0827	(0.900)	0.0321	(0.882)
Academic engagement in 8th Grade	0.267	(0.913)	0.272	(0.921)	0.252	(0.888)
Grit in 8th Grade	0.0442	(0.836)	0.0525	(0.842)	0.0183	(0.817)
School connectedness in 8th Grade	0.140	(0.900)	0.144	(0.906)	0.124	(0.882)
Study habits in 8th Grade	0.152	(0.889)	0.165	(0.900)	0.113	(0.853)
Absences in 8th Grade	8.606	(8.003)	7.972	(8.003)	10.629	(11.27)
GPA in 8th Grade	2.803	(0.781)	2.850	(0.772)	2.659	(0.791)
Days Suspended in 8th Grade	0.431	(1.812)	0.348	(1.546)	0.690	(2.442)
Incidents in 8th Grade	0.0636	(0.334)	0.0517	(0.286)	0.100	(0.451)
Demographics						
Female	0.502	(0.500)	0.512	(0.500)	0.469	(0.499)
Special education (IEP)	0.182	(0.386)	0.157	(0.364)	0.258	(0.438)
Free lunch	0.777	(0.416)	0.768	(0.422)	0.803	(0.397)
Reduced-price lunch	0.0753	(0.264)	0.0786	(0.269)	0.0650	(0.247)
White	0.0898	(0.286)	0.0951	(0.293)	0.0734	(0.261)
Black	0.403	(0.490)	0.374	(0.484)	0.491	(0.500)
Native	0.00169	(0.0411)	0.00163	(0.0403)	0.00190	(0.0435)
Asian/Pacific Islander	0.0327	(0.178)	0.0364	(0.187)	0.0215	(0.145)
Latinx	0.460	(0.498)	0.480	(0.500)	0.398	(0.489)
Multiracial	0.00147	(0.0383)	0.00149	(0.0386)	0.00138	(0.0371)
Long-term						
Ever arrested in school	0.041	(0.197)	0.039	(0.194)	0.048	(0.213)
HS Graduation	0.745	(0.436)	0.781	(0.414)	0.646	(0.478)
Enrolled in any college within 2 years	0.536	(0.499)	0.580	(0.494)	0.414	(0.493)
Enrolled in a 2 year college within 2 years	0.278	(0.448)	0.297	(0.457)	0.225	(0.418)
Enrolled in a 4 year college within 2 years	0.346	(0.476)	0.379	(0.485)	0.251	(0.434)
N	157630		119133		38497	

Notes: Survey completers are students who have 9^{th}-grade data for emotional health, academic engagement, grit, school connectedness, and study habits. As such, we report averages for some measures even among non-completers because many noncompleters are missing some data but not others.

Table S4: Psychometric Properties of SED measures (*as reported by the University of Chicago Consortium on School Research*): 2011 through 2013

Measure	School Year	Separation	Reliability	Item Infits	Item Outfits
Grit	2010-11	1.68	0.74	0.84, 0.76, 0.71, 1.24	0.85, 0.76, 0.71, 1.19
Social Skills	2010-11	1.69	0.74	1.08, 1.36, 1.41, 1.11	1.05, 1.33, 1.44, 1.15
Academic Effort	2010-11	1.74	0.75	0.85, 1.22, 1.1, 0.91	0.82, 1.17, 1.12, 0.94
Academic Engagement	2010-11	1.59	0.7	0.49, 0.56, 0.71, 0.56	0.49, 0.57, 0.72, 0.58
Belonging	2010-11	2.07	0.81	0.93, 1.02, 0.99, 0.96, 1.29	0.91, 0.97, 0.99, 0.93, 1.33
Grit	2011-12	1.54	0.7	0.8, 0.73, 0.68, 1.19	0.81, 0.57, 0.6, 0.42
Social Skills	2011-12	1.68	0.74	1.37, 1.36, 1.28, 1.06	1.68, 1.24, 1.18, 0.95
Academic Effort	2011-12	1.75	0.75	0.85, 1.22, 1.08, 0.92	0.82, 1.17, 1.1, 0.96
Academic Engagement	2011-12	1.56	0.71	0.54, 0.53, 0.47, 0.69	0.56, 0.55, 0.48, 0.71
Belonging	2011-12	2.13	0.82	0.98, 1.28, 0.91, 1.02, 0.97	0.97, 1.32, 0.89, 0.97, 0.94
Grit	2012-13	1.55	0.71	0.77, 0.69, 0.63, 1.13	0.79, 0.7, 0.63, 1.1
Social Skills	2012-13	1.67	0.74	1.3, 1.37, 1.23, 1.04	1.55, 1.25, 1.12, 0.94
Academic Effort	2012-13	1.77	0.76	0.86, 1.2, 1.13, 0.94	0.83, 1.15, 1.15, 0.97
Academic Engagement	2012-13	1.57	0.71	0.55, 0.54, 0.47, 0.69	0.57, 0.56, 0.48, 0.70
Belonging	2012-13	2.14	0.82	0.95, 1.28, 0.90, 1.03, 0.96	0.95, 1.31, 0.87, 0.98, 0.93

Notes. The reported statistics are from internal documentation at the University of Chicago Consortium on School Research where Rasch analysis was performed on individual survey items. All measures are anchored to 2010-11 step and item difficulties. Infit and outfit measures greater than 1 indicate underfit to the Rasch model and values lower than 1 indicate overfit. Generally, infit and outfit values in the range of 0.6-1.4 are considered reasonable for survey measures. Reliability represents individual reliability and includes extreme people. The patterns are very similar for years 2013 through 2018.

A Correlations Across Measures

Given that we have school impacts on several skill measures, it is helpful to see if they are related to each-other. To explore this, we report the correlations between the school impacts on the various skill measures in 9^{th} grade. These are reported in the lower panel of Table S5. The estimated school impacts (across the 133 schools) are all positive. That is, schools that improve one skill measure tend to improve the others. While all the correlations are positive, some variables are more closely related than others. For example, the correlation between math test score impacts and English test score impacts is 0.645. This is consistent with other studies. Interestingly, the correlation between test score impacts and impacts on the SED measures are reasonably large. The correlation between math value-added and the value-added on the individual surveys constructs are between 0.25 and 0.45. The correlations are similar for school impacts on English scores and the various surveys. To explore the extent to which schools tend to cluster in their impacts, we also conduct exploratory factor analysis. The factor ladings are reported in Table S6. The factors models suggests three distinct underlying factors. The first factor is most strongly related to the two test score impacts. The second factor is most strongly related to the survey measures relating to academic motivation and effort (grit, study habits, and academic engagement). The third factor is most strongly related to the two survey measures that relate to social well being (belonging, and social skills). While there is not complete separation of variables in the model, it clearly identified three district dimensions of school output.

Based on these results, we create a social index by combining the two social survey questions, we create a hard working index by combining the academic engagement questions, and we create a test score index by combining the two test scores. We also create an overall survey index that combines all the survey questions. To create each index we (1) standardize each measure, then we (2) compute the arithmetic mean across the included measures, and (3) standardize the combined index to be mean zero unit variance. The correlations across school impacts on the indexes are in the lower rows on Table S5. The two survey dimensions are highly correlated, but not perfectly so. The correlations between school impacts on the social index and the hard work index is 0.638. As we show, while there is much shared variation, there is explanatory power in the independent variation in each index. The correlation between school impacts on test scores and impacts on the summary survey indexes are between 0.4 and 0.45. This suggests that any single index could likely predict improved outcomes in an average sense. However, the policy relevant question is whether the school impacts on SED can provide additional information, and the extent to which school impacts on self-reported SED measures reflect real improvements.

Table S5: Correlations of School Value Added: Over time and Across Outcomes

	Correlations of Value-Added Within Outcomes Across Time (735 School-Year Observations)									
lag	Math Value Added	ELA Value Added	Grit Value Added	Interpersonal Skills Value Added	Academic Engage-ment Value Added	School Connect-edness Value Added	Study Habits Value Added	Test Scores Value Added	Social Value Added	Workhard Value Added
t+1	.404	.316	.26	.2	.217	.427	.174	.394	.356	.256
t+2	.266	.138	.104	.069	.165	.242	.164	.234	.17	.184
t+3	.136	.028	.048	.107	.078	.091	.012	.128	.077	.077
t+4	.146	.099	.108	.177	.109	.26	.184	.17	.227	.151

	Correlations of Average School-Level Value-Added Across Outcomes									
Math Value Added	1									
ELA Value Added	.645	1								
Grit Value Added	.311	.348	1							
Interpersonal Skills Value Added	.332	.261	.621	1						
Academic Engagement Value Added	.265	.11	.375	.374	1					
School Connectedness Value Added	.445	.434	.527	.668	.369	1				
Study Habits Value Added	.404	.455	.733	.504	.491	.554	1			
Test Scores Value Added	.917	.891	.353	.317	.202	.478	.462	1		
Social Value Added	.432	.392	.619	.882	.405	.939	.58	.446	1	
Workhard Value Added	.4	.359	.808	.593	.79	.578	.883	.408	.638	1

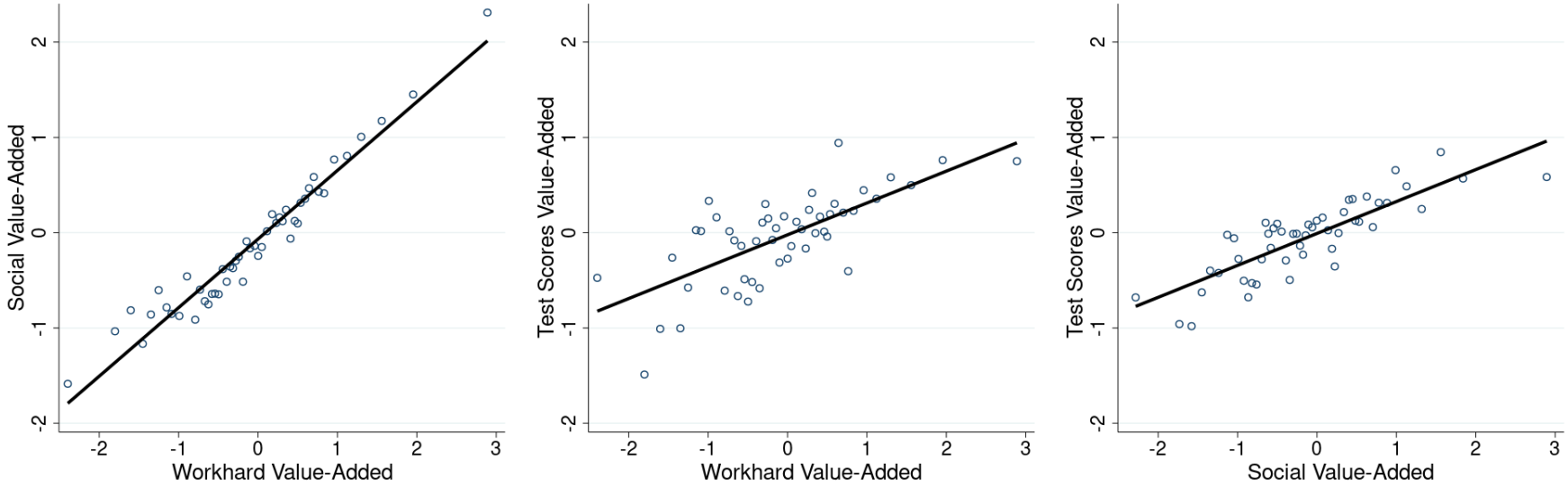
Notes: Restricted to school-year cells with at least 10 respondents

Table S6: Factor Loading for different Value- Added Measures

	Factor 1: *Test Scores*	Factor 2: *Work Hard*	Factor 3: *Social Well-Being*
Math Value-Added	0.6874	0.1889	0.2203
ELA Value-Added	0.7125	0.2231	0.1187
School Connectedness Value-Added	0.3211	0.3806	0.6155
Interpersonal Skills Value-Added	0.1344	0.4358	0.6494
Academic Engagement Value-Added	0.1789	0.5051	0.3417
Grit Value-Added	0.1424	0.7373	0.3513
Study Habits Value-Added	0.3151	0.7423	0.2515

Notes: These estimates are based on an exploratory factor analysis using the Bartlett Method. The rotated factor loadings are reported.

Figure S1. Binned Scatterplot of The Value-Added Measures



Notes: This figure presents binned scatterplots of the different value-added measures against each-other. There are 50 bins in each plot. These models do not include any controls.

B Stability of School Effects Over Time

The first test of whether schools have systematic impacts on survey measures is whether school impacts in one year are correlated with impacts in other years. As pointed out in Jackson (2014) in the presence of transitory shocks, standard analysis of variance may lead one to conclude that schools have systematic impacts when in fact they do not. As such, an alternate approach is to explore the extent to which school value-added is persistent over time. The basic logic is that if a school is able to systematically improve students' SED, it should be able to do so in multiple years. To test this, for each of the SED or test score measures in 9^{th} grade, we computed our annual value-added estimates from (4) and correlated them for the same schools over time. We report the correlations between a school's value-added in year t and in years $t+1$ through $t+4$ in the top panel of Table S5.

We start with the test score impacts. Looking at math scores, the correlation between a schools impact in adjacent years (i.e., years t and $t+1$) on math scores is 0.403. Specifically, a school that is above average at raising math scores in one year is likely to be above average in the following year. However, the relationship is far from one-for-one. Specifically, a school that was at the 98^{th} percentile of math score value-added in one year, would be expect to be at the 81^{th} percentile the following year. Looking at relationships two years apart, the correlation falls to 0.266. This means that a school that was at the 98^{th} percentile of math score value-added in one year, would be expect to be around the 70^{th} percentile two years later. Looking at more than three years out, the correlation stabilizes at around 0.14. This suggests that a school that was at the 98^{th} percentile of math score value-added in one year, would be expect to be around the 62^{nd} percentile three to five years later. This is clear evidence that school impacts do persist over time, but that the extent of the persistence is modest. It is important to note that school value-added is estimated with error so that these correlations are a lower bound on what we would observe if school impacts were perfectly measured (i.e., with no errors).

The basic pattern of persistent effects that are stronger for adjacent years is observed for all the SED and test score measures. For grit, the one-year correlation is 0.26. This persistence level is smaller than that for math, suggesting that school impacts on grit are nosier than those in math test scores. This one-year correlation is similar for the other survey measures other than school connectedness (which has a correlation of 0.427). To put these correlations into perspective, a correlation of 0.26 implies that a school that was at the 98^{th} percentile of grit, effort, academic engagement, or social skills value-added in one year, would be expect to be around the 70^{th} percentile the following later. As with the test score impacts, the persistence is lower with greater temporal distance. For all the survey measures, the correlation between impacts in year t and year $t+4$ is about 0.15. This implies that a school that was at the 98^{th} percentile of value-added for any of the survey measures in one year, would be expect to be around 64^{th} percentile three later. While these persistence effects are modest, they are statistically significantly different from zero – compelling evidence that schools do have systematic impacts on self-reported measures of socio-emotional development.

In sum, the results indicate that (a) both SED and test-score value-added in one year is predictive of value-added in the next, and (b) value-added estimates for temporally close years will be a better predictor of value added than years that are temporally distant. In essence, this reveals that school value-added (on both test scores and SED) has some persistence over time but does exhibit some "drift". We will exploit this fact when predicting a schools impact on hard or soft skills based on impact on other years.

Readers may notice that some of the correlations we document over time seem low relative to those documented for teacher value-added. This is driven by the fact that we include all school-year cells with 10 or more observations. If one were to restrict the sample to those school-year cells with more than 25 or even 50 students, the correlations are higher and are in line with others in the teacher value-added literature. See Table S7. Indeed, the correlations between year t and year t+4 are between 0.2 and 0.3 - similar to the stability of teacher value added documented in other studies.

Table S7: Correlations Between Value-Added Estimates Over Time: By Minimum Observations Per School-Year

lag	Math Value Added	ELA Value Added	Grit Value Added	Emotional Health Value Added	Academic Engage-ment Value Added	School Connect-edness Value Added	Study Habits Value Added	Test Scores Value Added	Social Value Added	Workhard Value Added
Restricted to schools with at least 10 respondents										
t+1	.404	.316	.26	.2	.217	.427	.174	.394	.356	.256
t+2	.266	.138	.104	.069	.165	.242	.164	.234	.17	.184
t+3	.136	.028	.048	.107	.078	.091	.012	.128	.077	.077
t+4	.146	.099	.108	.177	.109	.26	.184	.17	.227	.151
Restricted to schools with at least 25 respondents										
t+1	.415	.296	.313	.259	.263	.436	.24	.394	.386	.323
t+2	.29	.098	.136	.063	.232	.198	.244	.223	.144	.265
t+3	.169	.006	.053	.086	.074	.051	-.001	.137	.048	.05
t+4	.132	.063	.125	.268	.131	.214	.218	.146	.249	.192
Restricted to schools with at least 50 respondents										
t+1	.434	.358	.296	.282	.306	.423	.266	.432	.386	.332
t+2	.277	.147	.14	.073	.187	.205	.156	.254	.166	.174
t+3	.246	-.008	.127	.123	.029	.044	.058	.184	.078	.088
t+4	.205	.145	.163	.208	-.036	.187	.205	.233	.209	.137

C Testing For Selection

Because students are not randomly assigned to schools, there is a concern that our estimated value-addeds are related to unobserved predictors of outcomes so that our estimates are biased. While there is no way to prove that the value-addeds of the attended schools are unrelated to unobserved determinants of outcomes, we present several test to show that this is likely satisfied in our setting.

No Selection on Observables

First, to show that our school SED value-addeds are likely unbiased, we show that they are unrelated to observed determinants of student outcomes. That is, similar to a test for random assignment, we show that there is balance of covariates between high and low value-added schools. To show this, we estimate the following model by Ordinary Least Squares (OLS).

$$Z_{ijt} = \sum_{q \in Q} \pi_q \hat{\mu}_{jt,q} + \tau_t + \varepsilon_{ijt} \tag{5}$$

The parameter estimates of π_q provide a test of whether the observed covariate (Z_{ijt}) is correlated with the value-added on dimension q. If strong observable predictors of the outcomes are unrelated to our school value-added estimates (i.e. $\pi_q = 0$ for all covariates), then it is plausible that *unobservable* predictors are also unrelated to our value-added estimates so that our estimates are unbiased. We show evidence of this empirically in Table S1. If our estimated school value-addeds were correlated with student characteristics, then the coefficient on the value-added predictors would be significantly different from zero. We estimate this model with all three predictors across 33 pre-treatment student characteristics – resulting in 99 estimates. None of these point estimates is statistically significant at the 5 percent level. Remarkably, this is also true across the 99 models that have each value-added dimension individually – indicating no selection on observables. To summarize the result in Table S1, we also estimate effects on "predicted" measures and outcomes based on all of the observed covariates. We regress each outcome (graduate high school, enroll in college, etc.) or measure (test scores, work hard, and social) on all of the observed covariates to create predicted values. To avoid mechanical correlation, we form this prediction based on the regression from other years. This results in a leave-year-out predicted outcome or measure. We then regress these predicted outcomes and measures on the value-added estimates. The results are in Table S2. As expected, the value-addeds are jointly not significantly related to predicted outcomes. While this evidence supports a causal interpretation of our estimates, we also present tests of selection in unobserved dimensions below.

Attendance Boundary Instruments

Even though we show no evidence of selection on observables, one may worry about selection on unobservables. To address this, we construct instruments that remove the sorting bias that may exists when individuals chose to attend a school outside their zoned area. In Chicago, almost two-thirds of children attend schools other than their zoned school (Hing and Jenniver), so that this is a potential concern. To show that this does not bias our results, we propose an instrumental variables approach that instruments for the value-added of the school attended with the value-added of the residentialy assigned school. This approach eliminates all selection to non-zoned school that could have led to bias.

The first stage regressions are strong for all value-addeds – yielding first stage F-statistics above 20. The two-stage-least-squares (2SLS) regressions are reported in Table S8. Looking at work hard value-added (middle panel), a comparison of columns 1 and 2 reveals that the OLS and 2SLS models for work hard in 9th grade are similar and not statistically distinguishable from one-another. Looking at columns 4 and 5, and then 7 and 8, reveals that the OLS and 2SLS models of work hard value-added are also very similar for high school completion and college-going. Owing to much larger standard errors in the IV model, the point estimate on college-going is not significant, but it is very similar to the OLS estimate. Note however that the effect on high school graduation is significant at the five percent level in both the OLS the 2SLS models. The lower panel present a very similar pattern for social well-being value-added.[10] In sum, our SED value-added measures do not appear to be biased by selection on unobservables. These 2SLS estimates will only be biased if those families that attend the zoned schools tend to self-select into neighborhoods along unobserved dimensions that are correlated with school value-added. We address this possibility.

[10]The top panel reports impacts for test score value-added. Unlike the SED value-addeds, these models are less consistent across specifications. However, test score value-added is not the focus of our study.

Sibling Comparisons

To account for the possibility that families may select into neighborhoods in ways that would lead to bias in our 2SLS approach, we also estimate models that rely on within family comparisons. For a small subset of the data we are able to identify siblings. That is, we can identify siblings in the data after 2015. As such, for families that have more than one sibling who were in CPS after 2015 we can make within-family comparisons. We were able to identify 13,150 families in which more than one sibling is observed in 9^{th} grade. Of these that have multiple children old enough to have graduated from high school we have 3822 such families. For those old enough to have enrolled in college, this number falls to 1581 families. Because we cannot identify *all* siblings prior to 2015, these data are imperfect and incomplete. However, if we are able to find similar effects in this small sub-sample as in the broader sample, it would be compelling evidence that our estimates are not biased by family selection to neighborhoods. We can remove any correlation with potentially confounding family characteristics by comparing students from the same family who attended different schools. This is achieved by adding a family fixed effect to our main model in equation (4). The within-family estimates are presented in Table S8. The sibling models are presented in the third, sixth, and ninth column of the table. As one can see, in the lower two panels, the estimated impacts of SED value-added are robust to the inclusion of the family fixed effects. That is, while the standard errors are much larger in the family fixed effects models, the point estimates for impacts on both work hard and social well-being value added on high school completion and college going are very similar to the OLS models.[11] This indicates that selection of families does not drive the estimates.

Taken together, we show that (a) our value-added estimates are unrelated to all observed covariates, (b) our estimates are not driven by selection to schools outside one's attended zone, and (c) our estimates are not biased by certain kinds of families sending their children to different schools. If our results were driven by selection to schools across families, it would bias our IV results but not our sibling results. If our results were driven by selection to schools within families, it would bias our sibling results but not our 2SLS results. If there were selection (either within or across families) one would expect that strong predictors of outcomes would be related to our estimated value-added– but this is not the case. While none of these tests is dispositive in isolation, together they are compelling evidence that our estimated school impacts, and the main results, reflect true causal impacts and are not driven by any selection bias.

[11] The top panel reports impacts for test score value-added. Unlike the SEL value-added, these models are less consistent across specifications. However, test score value-added is not the focus of our study.

Table S8: Selection on Unobservables Models

	1	2	3	4	5	6	7	8	9
		Test Scores			High School Graduation			Enroll in College	
Test Score Value-Added	0.061	0.0422	0.0227	0.0105	0.0073	0.00450	0.0151	0.00344	-0.00281
	(0.0111)	(0.00526)	(0.0122)	(0.00243)	(0.00287)	(0.00460)	(0.00457)	(0.00393)	(0.00527)
Observations	101,984	98,050	16,361	81,885	78,318	8,160	55,378	52,439	3,370
OLS	X			X			X		
School Assignment IV		X			X			X	
First-Stage F-statistic		1922			2200			1782	
Sibling Fixed Effects			X			X			X
Number of Families			7596			3822			1581

	1	2	3	4	5	6	7	8	9
		Work Hard			High School Graduation			Enroll in College	
Work Hard Value-Added	0.064	0.0788	0.0248	0.0159	0.0256	0.0165	0.0197	0.0171	0.0135
	(0.00892)	(0.0107)	(0.0133)	(0.00375)	(0.00737)	(0.00617)	(0.00595)	(0.0102)	(0.0153)
Observations	124,051	120,088	28,620	81,725	78,182	8,141	55,218	52,303	3,358
OLS	X			X			X		
School Assignment IV		X			X			X	
First-Stage F-statistic		621.3			793.4			650.1	
Sibling Fixed Effects			X			X			X
Number of Families			13105			3813			1575

	1	2	3	4	5	6	7	8	9
		Social			High School Graduation			Enroll in College	
Social Value-Added	0.0905	0.0891	0.0607	0.0156	0.0352	0.0106	0.0172	0.0369	0.0111
	(0.00801)	(0.0122)	(0.0142)	(0.00358)	(0.00777)	(0.00566)	(0.00528)	(0.0101)	(0.00976)
Observations	124,248	120,270	28,723	81,725	78,182	8,141	55,218	52,303	3,358
OLS	X			X			X		
School Assignment IV		X			X			X	
First-Stage F-statistic		568.4			871.4			715	
Sibling Fixed Effects			X			X			X
Number of Families			13150			3813			1575

Robust standard errors adjusted for clustering at the school level in parentheses

Notes: Results are based on regression of each outcome on out-of-sample Value Added of Social, Workhard, and Test Scores Value Added. All models include individual demographic controls (race / ethnicity, free and reduced price lunch, and gender and the socioeconomic status of the student's census block), 8^{th} grade lags (math and ELA test scores, survey measures, absences, and discipline), and school-level averages for the all demographics and lagged measures, as well as year fixed effects. We also include the socio-economic status of the student census block proxied by average occupation status and education levels. The instrumental-variables model uses the value-added of each student's zoned school as an instrument for the value-added of the school they attend. The sibling model includes a sibling fixed effect.

D Additional Tables

Table S9: Effects on 9^{th} Grade GPA

	1	2	3	4	5
	Core GPA in 9^{th} Grade (obs. 122,714)				
Social Value-Added	0.0353			0.00805	0.00610
	(0.0188)			(0.0171)	(0.0170)
Work Hard Value-Added		0.0452		0.039	0.0367
		(0.0221)		(0.0233)	(0.0228)
Test Scores Value-Added			0.0235		0.0155
			(0.0109)		(0.00954)
SD of Predicted School Effect	0.0350	0.0452	0.0262	0.0451	0.0495
P-value from F-test of Work Hard and Social				0.1285	0.1622

Robust standard errors adjusted for clustering at the school level in parentheses.

Notes: The dependent variable is 9^{th}-grade Grade-Point Average (GPA) in core courses (Math, English, Social Studies and Science) computed from CPS course files. Results are based on regression of 9^{th}-grade GPA on out-of-sample Value Added of Social, Workhard, and Test Scores Value Added. All models include individual demographic controls (race / ethnicity, free and reduced price lunch, and gender), 8^{th} grade lags (math and ELA test scores, survey measures, absences, and discipline), and school-level averages for all the demographics and lagged measures, as well as year fixed effects. We also include the socio-economic status of the student census block proxied by average occupation status and education levels. For each model, we estimate the standard deviation of the predicted school impacts based on the value-added measures included in the model (i.e. $\hat{\sigma}_{\hat{F}}$). In columns 4 and 5, we report the p-value associated with the test that the coefficients on Work hard value added and Social value-added are jointly equal to zero. Missing 8^{th} grade measures were imputed using 7^{th} grade measures and demographic characteristics.