PACE
Policy Analysis for California Education

WORKING PAPER

Stability of School Contributions to Student Social-Emotional Learning Gains

School value-added models are increasingly used to measure schools' contributions to student success. At the same time, policymakers and researchers agree that schools should support students' social-emotional learning (SEL) as well as academic development. Yet, the evidence regarding whether schools can influence SEL and whether statistical growth models can appropriately measure this influence is limited. Recent work shows meaningful differences across schools in changes in SEL scores by grade (Loeb, Christian, Hough, Meyer, Rice, & West, 2019), but whether these differences represent the effects of schools is still unclear. The current paper builds upon this earlier work by examining the stability of the estimated school-by-grade effects on SEL across two years, using a large-scale SEL survey administered in California's CORE districts. We find that correlations among school effects in the same grades across different years are positive, but they are lower than those for math and English Language Arts (ELA). Schools in the top or the bottom of the school effect distribution are more persistent in their impacts across years than those in the middle of the distribution. Overall, the results provide evidence that these school effects measure real contributions to SEL. However, the low stability of effects from one year to the next draw into question whether including these school value-added measures of self-reported SEL in school performance frameworks and systems would be beneficial.
.

Hans Fricke
Policy Analysis for California Education

Susanna Loeb
Brown University

Robert Meyer
Education Analytics

Andrew Rice
Education Analytics

Libby Pier
Education Analytics

VERSION: May 2019

Acknowledgements

This paper was produced as part of the CORE-PACE Research Partnership, which is focused on producing research that informs continuous improvement in the CORE districts (Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, Sacramento City, San Francisco, and Santa Ana unified school districts) and policy and practice in California and beyond. We thank Michael Christian, Nichole Webster, Tanner O'Brien, and Jordan Mader for their contributions to the analyses in this paper. The study benefited from the feedback of two anonymous reviewers as part of the PACE review process. Finally, we thank the CORE Districts for partnering with us in this research, providing the data for this study, and giving feedback on earlier drafts of this study. PACE working papers are circulated for discussion and comment purposes and have not undergone the peer-review process that accompanies official PACE publications.

Introduction

In recent decades, federal legislation such as the Every Student Succeeds Act (ESSA) and the early No Child Left Behind Act (NCLB) have accelerated the development and use of school performance measurement systems in districts and states across the U.S. The central purpose of these systems is to monitor how much schools contribute to student learning and success and to gauge schools' progress in improving student outcomes over time. The increasing focus on improvement, or growth, over time—rather than merely proficiency or attainment—is evident in the emergence of value-added models for measuring growth; currently, 48 states and the District of Columbia use such models to measure school performance (Data Quality Campaign, 2019). The strength of these models is that they model student outcomes as a function of student outcomes in the previous year and demographic characteristics, thereby accounting for differences in the student population that might bias estimates of schools' progress in improving students' outcomes.

At the same time, policy makers and researchers are increasingly interested in expanding the breadth of measures used to assess student success. In particular, social-emotional learning (SEL), sometimes called *non-cognitive skills*, has emerged as a priority for schools to consider as part of their definition of student success, in part because SEL has been shown to predict a wide range of outcomes in school and later in life (Heckman & Rubenstein, 2001; Almlund, Duckworth, Heckman, & Kautz, 2011; Heckman, Humphries, & Kautz, 2014; Deming, 2017). Moreover, these skills appear to be more malleable than more traditional measures of student success, such as standardized assessments of math and English Language Arts (ELA) (Cunha & Heckman, 2008; Dee & West, 2011; Heckman & Kautz, 2013). Furthermore, a growing body of research points to experiences in schools as one factor that can influence students' SEL. Studies suggests that schools can influence student SEL directly (Allensworth & Easton, 2007; Durlak, Dymnicki, Taylor, Weissberg, & Schellinger, 2011) and indirectly through the improvement of school culture and climate and the promotion of positive relationships within schools (Battistich, Schaps, & Wilson, 2004; Berkowitz, Moore, Astor, & Benbenishty, 2016; Blum, Libbey, Bishop, & Bishop, 2004; Hamre & Pianta, 2006; Jennings & Greenberg, 2009; McCormick, Cappella, O'Connor, & McClowry, 2015). Additionally, school-based interventions targeting SEL have proven effective for improving student academic achievement (Durlak et al., 2011).

Despite these recent findings, research establishing the degree to which schools contribute to students' social-emotional development, and whether growth models are suitable to measure this contribution, is limited. Our prior work (Loeb et al., 2019) aimed to fill this gap in the literature by applying school-by-grade level value-added models to students' SEL. We used a large-scale survey of students' self-reported SEL administered in California's CORE Districts, a consortium of eight large urban school districts. We provided estimates of school-by-grade effects (henceforth referred to as school effects) on four domains of SEL assessed by the survey: growth mindset, self-management, self-efficacy, and social awareness. We used models similar to those that have been used for estimating school effects on achievement and found that the variation of these effects was comparable to those on standardized assessment scores in math and ELA, which suggests that schools do differ measurably in how much their students' SEL measures change over a year. However, the results also showed that student scores in the prior year and their demographic characteristics explained substantially less variation in student self-reported SEL than they did for math and ELA scores. One plausible explanation for the lack of predictive power is that the SEL measures may have measurement error not captured by adjustments for sampling error. In this case, some of the estimated variation across schools might have resulted from noise in the underlying measure.

Building on this line of inquiry, the current paper assesses how stable school effects on SEL are over time. To answer this question, we use the same methodology as Loeb et al. (2019) to create growth measures and incorporate one additional year of data to calculate school value-added models for growth mindset, self-management, self-efficacy, and social awareness separately for two years (2015-16 and 2016-17). We then compare the results of these two models across years. Specifically, we calculate the correlation of school effects in the same grade in different years, and the correlation of school effects following cohorts in adjacent grades in different years. Thus, the results of this study aim to shed light on whether schools that appear to effectively support their students' social-emotional development in one grade in one year appear do so again for the next cohort, as well as whether schools that appear to contribute to one cohort's SEL in one year appear to further foster SEL for the same cohort in the next year or whether the cohort bounces back to closer to where it was initially. Finally, this study extends our prior work in Loeb et al. (2019) by providing estimates of the variance of school effects on SEL for students in high school in addition to students in grades four through eight.

Much of the literature on value-added models has focused on estimates of teacher-level effects (e.g. Rivkin, Hanushek, & Kain, 2005; Chetty, Friedman, & Rockoff, 2014). Recently, studies have applied teacher-level value-added models to estimate teacher effects on students' SEL and other non-cognitive outcomes (Jennings & DiPrete, 2010; Gershenson, 2016; Ladd & Sorensen, 2017). Specifically, teachers have been shown to influence academic motivation (Ruzek, Domina, Conly, Duncan, & Karabenick, 2015), self-efficacy and happiness (Blazar & Kraft, 2017), and suspensions and attendance (Jackson, 2018; Liu & Loeb, 2018). In contrast, although research on school-level effects for traditional academic measures, such as standardized test scores in math and ELA, is robust (Deming, 2014; Grissom, Kalogrides, & Loeb, 2015; Chiang, Lipscomb, & Gill, 2016; Angrist, Hull, Pathak, & Walters, 2016, 2017), the literature typically has not applied these models to school-level effects on students' SEL and non-cognitive outcomes.[1]

Evidence on the stability of school value-added measures over time is limited and focuses on measures of academic performance as assessed by standardized assessments. Early research about the persistence of school quality over time shows mixed conclusions (Teddlie & Reynolds, 2000). These differences appear to be driven by differences in modeling approaches: the more that models control for endogenous factors of student composition, the lower the persistence of the respective school quality measures (Gray, Goldstein, & Thomas, 2001; Thomas, Peng, & Gray, 2007; Dumay, Coe, & Anumendem, 2014, Marks, 2015). Similarly, evidence on teacher-level

[1] Several studies have addressed estimation issues of school growth models (Raudenbush & Willms, 1995; Meyer, 1997; Tekwe et al., 2004; Reardon & Raudenbush, 2009; Ehlert, Koedel, Parsons, & Podgursky, 2016).

effects also points to mixed results. In a recent literature review, Koedel, Mihaly, and Rockoff (2015) report that year-to-year correlation in estimated teacher value-added effects range from 0.18 to 0.64. Thus, the literature on the stability of school value-added measures over time, particularly for new kinds of outcomes measured by new kinds of assessments, is nascent.

The results in this paper suggest that schools that appear to improve students' SEL in one year may not necessarily do so again in the next, based on school value-added measures on self-reported SEL. Overall, we find that the correlations among school effects in the same grade across different years for growth mindset, self-management, self-efficacy, and social awareness are positive, but relatively small, ranging from 0.03 to 0.54, with the exception of self-efficacy in grade nine (0.74). Overall, these patterns provide further evidence that estimates of school effects on students' self-reported SEL measure real contributions to student SEL but are not very stable from one year to the next. Nonetheless, we show that a significant number of schools stand out by being consistently in the top or bottom of the school effect distribution in both years.

The lack of stability of school effects is consistent with two explanations. First, schools may affect students but in idiosyncratic ways that do not persist from year to year. Second and more likely, omitted factors in the growth model and measurement error of the SEL measures may inflate the variance of estimated school effects such that the estimates capture more noise than true effects of schools. In either case, the low stability of school effects raises concerns about the degree to which school value-added models using self-reported SEL survey measures can distinguish the SEL contributions of most schools, though a significant number of schools may be consistently impactful enough to distinguish. As the analyses presented in this paper are the first of their kind, they highlight the need for additional research into the use of any such measures to gauge or report on school performance.

Data

We rely on data from five of the eight CORE districts (Fresno, Long Beach, Los Angeles, San Francisco, and Santa Ana). The CORE districts are a consortium of large urban school districts together serving more than one million students, approximately 20 percent of all students in California. Since 2014, the CORE districts have administered survey-based measures of students' SEL each spring, asking students to rate themselves on questions related to growth mindset, self-management, self-efficacy, and social awareness. Students choose one of up to five available responses best describing their agreement or their participation in an activity or experience. We use survey responses in school years 2014-15, 2015-16, and 2016-17.

The four SEL constructs are defined as follows: *Growth mindset* is the belief that one's abilities can grow with effort. Students with a growth mindset see effort as necessary for success, embrace challenges, learn from criticism, and persist in the face of setbacks (Dweck, 2006). *Self-efficacy* is the belief in one's own ability to succeed in achieving an outcome or reaching a goal. Self-efficacy reflects confidence in the ability to exert control over one's motivation, behavior, and environment (Bandura, 1997). *Self-management* is the ability to regulate one's emotions, thoughts, and behaviors effectively in different situations. This includes managing stress, delaying gratification, motivating oneself, and setting and working toward personal and academic goals (CASEL, 2005). *Social awareness* is the ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources (CASEL, 2005).

The CORE districts began developing the survey in 2013 as part of an alternative school accountability system proposed as part of a waiver from the then-mandated NCLB. CORE convened SEL experts and stakeholders to identify research-based SEL constructs that are meaningful to stakeholders, that are malleable, and that can be measured. The survey was pilot tested in 18 schools in the 2013-14 school year and then administered in all districts beginning in the 2014-15 school year. West, Buckley, Krachman, and Bookman (2018) provide a more detailed description of the survey development process. With the passage of ESSA in 2015, the CORE districts opted not to include the SEL survey as a school accountability measure, but rather to report school-wide descriptive measures of SEL in their reporting dashboard in order to inform continuous improvement efforts by local educators and administrators.

In addition to the SEL survey, the data in this study includes math and ELA scores from the Smarter Balanced Assessment Consortium (SBAC). All students in California take this standardized, computer-adaptive test in the spring of grades three through eight and grade 11. Finally, the study uses demographic student variables, including indicators for economic disadvantage, special education, English learner (EL) status, foster youth, homelessness, and race/ethnicity.

In total, we estimate school effects for six measures: math SBAC scores, ELA SBAC scores, growth mindset, self-management, self-efficacy, and social awareness. We estimate these models separately for outcome measures in a given grade in school years 2015-16 and 2016-17. In the models, we control for pretest scores in all six outcomes (just the four SEL outcomes in grades 10 and 11) in 2014-15 and 2015-16. We include students in the sample who answered at least one of the items corresponding to the SEL construct used as the outcome or who have the corresponding SBAC score. Moreover, we include only students who answered at least one of the items for each SEL measure used as a pre-test control and who have both SBAC scores. Finally, we restrict the sample to students who had demographic information and who were continuously enrolled at a specific school. The final sample includes between 19,944 and 49,081 students and between 167 and 721 schools for each construct, year, and grade (see Table 1, Panel A and B). Although the overall sample sizes are very similar across outcomes, samples for SBAC scores are slightly bigger than those for the SEL constructs due to non-response in the survey. Sample sizes decrease in higher grades.

Table 1. Descriptive Statistics

Panel A: Number of students (samples for all four outcomes)

Year	Outcome	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
2016	Math	41633	40506	37943	38454				
	ELA	41649	40541	37969	38531				
	Growth Mindset	37455	33553	31344	31675	25296	26967	22737	19996
	Self-Efficacy	37516	33564	31366	31702	25299	26993	22736	20003
	Self-Management	37560	33637	31443	31760	25349	27017	22784	20017
	Social Awareness	37267	33441	31203	31598	25213	26907	22669	19944
2017	Math	49081	42444	38034	37789				
	ELA	49070	42418	38001	37794				
	Growth Mindset	44129	36473	31715	31776	24701	23995	22837	19978
	Self-Efficacy	44102	36366	31596	31678	24655	23941	22778	19946
	Self-Management	44574	36972	32292	32218	25054	24330	23103	20187
	Social Awareness	44715	37173	32456	32365	25160	24424	23177	20270

Panel B: Number of schools (samples for all four outcomes)

		Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12
2016	Math	721	340	196	197				
	ELA	721	340	196	197				
	Growth Mindset	706	330	193	194	167	209	210	203
	Self-Efficacy	706	330	193	194	167	209	210	203
	Self-Management	707	330	193	194	167	209	209	204
	Social Awareness	707	330	193	194	167	209	211	204
2017	Math	615	340	192	192				
	ELA	615	340	192	191				
	Growth Mindset	615	334	187	189	168	209	213	192
	Self-Efficacy	615	334	187	189	168	209	213	191
	Self-Management	615	334	187	189	168	209	213	192
	Social Awareness	615	334	187	188	168	209	213	192

Panel C: Demographics (proportions of students by characteristic, growth-mindset outcome sample)

	Grade	% ELL	% SWD	% Econ. Disadv.	% Home-less	% Foster	% Latinx	% White	% African American	% Asian
2016	5	18.16	11.29	77.56	1.85	0.73	73.44	10.25	7.36	5.00
	6	14.40	10.92	78.63	2.07	0.62	71.06	9.59	6.83	7.84
	7	11.59	10.23	77.02	2.35	0.48	69.79	9.99	6.36	8.59
	8	9.87	9.85	76.88	2.58	0.47	70.44	10.12	6.58	8.13
	9	9.99	9.74	78.98	2.16	0.38	72.34	8.20	6.26	8.36
	10	11.58	9.30	76.95	2.49	0.47	71.89	7.87	6.80	8.63
	11	7.90	8.80	75.41	1.92	0.32	69.79	8.15	7.15	9.50
	12	5.21	7.21	77.06	1.84	0.29	70.02	7.95	6.46	9.93
2017	5	16.87	10.32	79.31	3.66	0.46	71.65	10.02	6.9	7.10
	6	13.58	10.52	78.22	3.65	0.35	70.82	9.82	6.13	8.33
	7	10.33	9.86	77.07	3.41	0.32	69.42	9.81	6.53	9.06
	8	8.69	9.80	76.08	3.90	0.34	67.95	10.36	6.66	9.54
	9	8.62	9.14	79.8	2.77	0.31	72.58	8.13	6.31	8.19
	10	8.58	9.53	77.7	3.8	0.29	70.99	8.41	6.41	8.72
	11	8.47	8.57	76.51	4.19	0.28	70.62	8.66	6.26	9.10
	12	5.49	7.56	76.43	3.93	0.28	70.58	8.33	6.48	8.94

Table 1, Panel C describes the student population in participating districts. For simplicity, we show characteristics in the growth mindset sample. Characteristics in the samples for other outcomes are very similar. Seventy to 74 percent of students are Latinx, approximately seven percent are black, and about five to nine percent are Asian. In grade five, around 17 percent of students are classified as ELL, whereas this drops to around five percent in grade twelve. Around 11 percent of students require special education in grade five and around seven percent in grade twelve. Between 75 and 79 percent of students are economically disadvantaged (i.e., qualify for free or reduced-price lunch or have parents without a college degree). In the 2016 sample, around two percent of students are homeless; this number doubled in the 2017 sample. Less than one percent of students are in foster homes.

Instead of using raw survey responses, we create scale scores for each SEL construct. To do so, we use a generalized partial credit model (GPCM; Muraki, 1992) to aggregate all items of the corresponding construct. The GPCM is an extension of the partial credit model (Masters, 1982) and can incorporate responses to items on a multipoint scale. The GPCM weights items higher that better distinguish students with different construct-specific competencies and accounts for missing values in items.

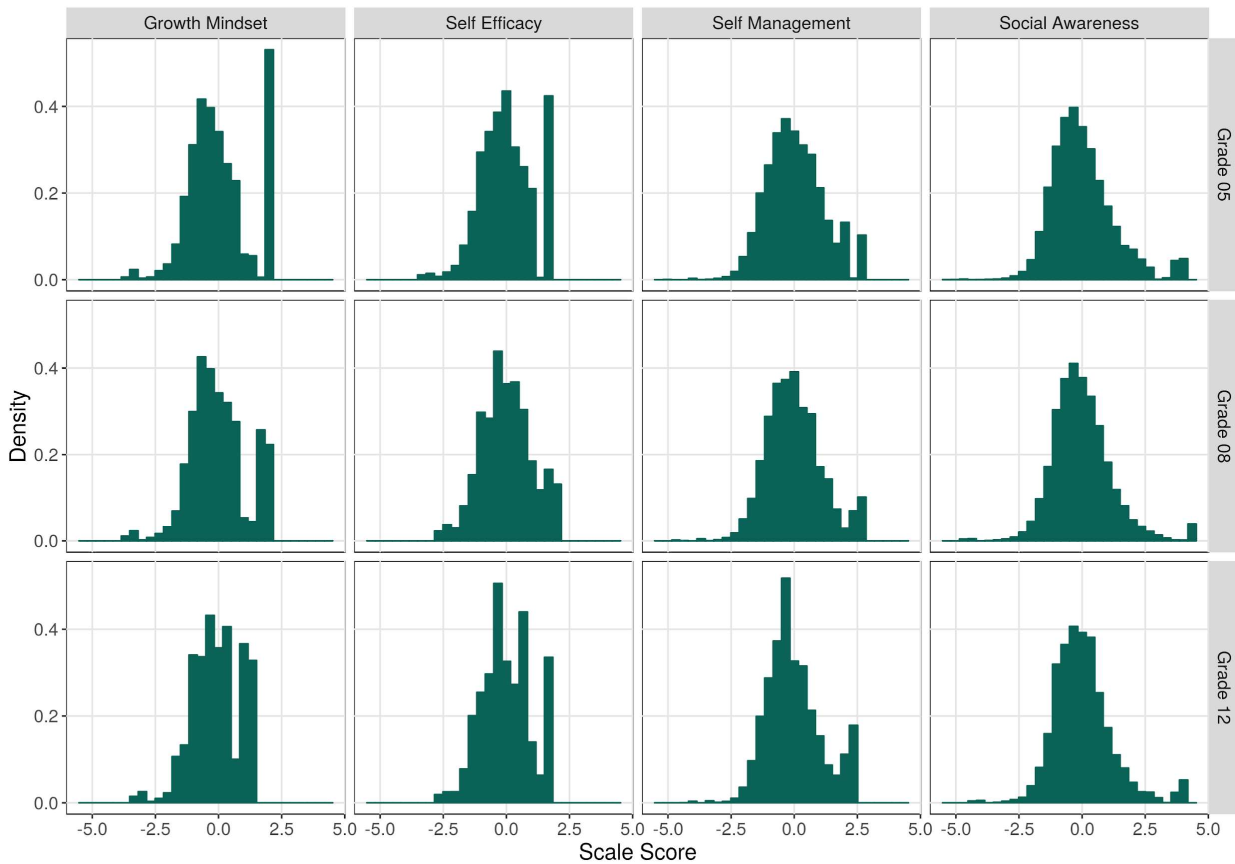
Previous research using the SEL data from the CORE districts has provided evidence for the validity and reliability of the survey measures and their scales (see Gehlbach and Hough, 2018, for an overview). Meyer, Wang, and Rice (2018) report that survey responses within each of the four SEL domains have high internal scale reliability for most grades,[2] that items across domains clearly load on distinct factors in exploratory factor analysis, and that students in different subgroups do not answer items differentially within the survey in differential item response analysis. In addition, West, Buckley, et al. (2018) find that the SEL constructs are predictive of several measures of student achievement such as GPA, ELA and math test scores, attendance, and suspensions. West, Buckley, et al. (2018) also provide evidence against reference bias, as the overall correlation between the SEL measures and academic measures exceeded the within-school correlation; if students rated themselves against their peers, the opposite should be true. Similarly, Claro and Loeb (2019a, 2019b) show that students' reports of growth mindset and self-management predict not only achievement levels, but also learning gains in math and ELA in the following year. Finally, West, Pier, et al. (2018) demonstrate that trends in these constructs over time as students move through school--and the differences of these trends among student subgroups--were largely consistent with findings in prior research. For instance, girls report higher levels of self-management and social awareness than boys, and socio-economically disadvantaged students report consistently lower SEL than their non-disadvantaged counterparts.

One potential concern with the survey measures is score bunching. If, for example, many students choose the highest possible answer (i.e., score of 5 out of 5), the scores will not provide evidence of variation among that top group. Relative to the use of raw means, GPCM scale scores

[2] Only the reliability within the growth mindset domain was below 0.7 for grades below grades seven, which is likely due to the negative wording of the growth mindset items.

help alleviate this concern by down-weighting items with a high proportion of a particular response option. Figure 1 shows the distribution of the GPCM scale scores for the four SEL constructs in in grades five, eight, and twelve (distributions are highly similar for other grades). Although there are notable spikes in the right tail of the distribution, the scores are overall smoothly distributed. Furthermore, the extent of the ceiling effects in the measures may not be a problem for the estimation of school-level effects (Koedel & Betts, 2010).[3]

Figure 1. Histogram of distribution of SEL scale (IRT theta) scores, grades 5, 8, and 12



Between- and Within-School Variance and Across-Year Covariance

We first explore the variance of the SEL constructs and SBAC scores themselves. We decompose the variance of the SEL constructs and SBAC scores, as well as their correlation from one year to the next, into a within-school and an across-school component. To do so, we estimate the following seemingly-unrelated-regressions (SUR) model:

[3] We also investigated the distribution of student-level growth scores across grades and constructs and found no evidence of ceiling or floor effects. See Figure A.1 in the Appendix.

$$Y_{cijt} = \mu_{cjt} + \eta_{cijt} \quad (1)$$

$$Y_{cijt-1} = \mu_{cjt-1} + \eta_{cijt-1} \quad (2)$$

where *j* is the school attended by student *i* in year *t*; Y_{cijt} is score in outcome *c* of student *i* in year t; μ_c is the component of the variance of outcome *c* in year *t* that is across year-*t* schools; and η_{cijt} is the component of the variance of construct *c* in year *t* that is within year-*t* schools. The school attended in year *t* is used to decompose the variance in both year *t* and year *t*-1 to ensure comparability to our school growth model. We report estimates of the variances of μ_{cjt}, μ_{cjt-1}, η_{cijt}, and η_{cijt-1}, and of the covariances between μ_{cjt} and μ_{cjt-1} and between η_{cijt} and η_{cijt-1}.

Table 2 presents the across-school and within-school variance of the SBAC scores in math, SBAC scores in ELA, and the four SEL scale scores for grades five, eight, and twelve for 2015-16 and 2016-17 (patterns in other grades are similar). The across-school component of the SEL scale score variances is smaller than the across-school component of variance in the SBAC scores. For example, in grade five in 2015-16, the across-school component represents four percent of the variance in social awareness, compared to 22 percent of the variance in the math score. Moreover, the across-school variance component appears to be decreasing in higher grades. For example, in grade 12, the across-school component represents only one percent of the variance in social awareness. This pattern is present in both years of data. The smaller across-school variation in the SEL measures could arise from greater measurement error or from a smaller school effect.

Table 2 also shows the across-school and within-school correlations between current (i.e., outcome) and lagged (i.e., pretest) scale scores. In both years, the year-to-year correlation in the SEL scale scores is lower than for the SBAC scores for both the across-school and within-school components. This pattern demonstrates that student-level SEL outcomes, as measured by the survey, have lower persistence over time than the academic measures.

Table 2: Across-school and within-school-across-student components of variance and year-to-year correlation in scale scores in academic subjects and SEL constructs

	Variance of scale scores, 2015-16		Correlation of scale scores, 2014-15 to 2015-16		Variance of scale scores, 2016-17		Correlation of scale scores, 2015-16 to 2016-17	
	Across-school	Within-school	Across-school	Within-school	Across-school	Within-school	Across-school	Within-school
Grade 5								
English language arts	0.21	0.79	0.94	0.8	0.2	0.8	0.94	0.81
Mathematics	0.22	0.78	0.93	0.83	0.22	0.78	0.92	0.83
Growth mindset	0.11	0.89	0.59	0.3	0.1	0.9	0.59	0.29
Self-efficacy	0.05	0.95	0.53	0.42	0.04	0.96	0.59	0.38
Self-management	0.07	0.93	0.78	0.5	0.04	0.96	0.67	0.43
Social awareness	0.04	0.96	0.54	0.41	0.04	0.96	0.45	0.35
Grade 8								
English language arts	0.16	0.84	0.94	0.81	0.18	0.82	0.95	0.83
Mathematics	0.18	0.82	0.95	0.82	0.2	0.8	0.96	0.86
Growth mindset	0.03	0.97	0.84	0.42	0.04	0.96	0.71	0.43
Self-efficacy	0.03	0.97	0.76	0.52	0.03	0.97	0.76	0.5
Self-management	0.05	0.95	0.86	0.52	0.03	0.97	0.71	0.49
Social awareness	0.04	0.96	0.82	0.47	0.05	0.95	0.59	0.44
Grade 12								
Growth mindset	0.02	0.98	0.67	0.46	0.01	0.99	0.56	0.46
Self-efficacy	0.02	0.98	0.59	0.51	0.03	0.97	0.6	0.5
Self-management	0.03	0.97	0.65	0.45	0.02	0.98	0.36	0.41
Social awareness	0.01	0.99	0.59	0.46	0.04	0.96	0.47	0.45

Growth Measures

We estimate schools' contributions to student growth in SEL, math, and ELA scores separately by grade and separately by year (i.e., 2015-16 and 2016-17) using the following regression model:

$$y_{cijt} = \xi_c + y_{Cijt-1}\lambda_c + X_{ijt}\beta_c + \alpha_{cjt} + \varepsilon_{cijt} \qquad (3)$$

where school *j* is the school attended by student *i* in year t; y_{cijt} is the outcome measure *c* for student *i* in school *j* in year t; α_{cjt} is the impact of school *j* on growth in construct *c* in year t; ε_{cijt} is a student error term; and λ_c and β_c are conformable coefficient vectors. X_{ijt} is a vector of demographic characteristics of student *i* in year t including indicators for economic disadvantage, special education, English learner (EL) status, foster youth, homelessness, and race/ethnicity. y_{Cijt-1} is a vector of all outcome measures available for student *i* in year *t*-1. Depending on the grade, this vector includes either all four SEL constructs and math and ELA scores (grades five through eight) or just the four SEL constructs (grade five through twelve).[4]

The model estimates the effect of a school on a student's outcome in year t while taking into account the level at which she starts the school year in year t-1; in other words, it estimates the school effect on student growth. The model accounts for any differences in the student

[4] McCaffrey, Lockwood, Koretz, Louis, and Hamilton (2004) have referred to this model specification as the covariate adjustment model, while Guarino, Reckase, and Wooldridge (2015) have called it the dynamic ordinary least squares model.

population among schools by controlling for lagged outcomes (student SEL scores and academic test scores in year t-1*)*, as well as demographic characteristics. The lagged outcomes are meant to capture all factors that have influenced student outcomes up to year *t,* while the demographic characteristics are meant to capture all non-school factors that may differ systematically between schools in year t. Under the assumption that all relevant factors that vary across schools are controlled for, α_{cjt} measures the contribution of school *j* to student growth.[5]

Note that, in contrast to academic measures, SEL is not necessarily expected to continuously increase as students move through school. West, Pier et al. (2018) show that, with the exception of growth mindset, the CORE SEL measures decrease after Grade 6. Therefore, when we talk about school contributions to growth in student SEL, we mean school contributions to change in student SEL relative to these trends. In other words, a school's contribution could be "high" if its students' SEL decreases less than would be expected compared to similar students in other schools, or a contribution could be "average" if its students' SEL decreased the same amount as would be expected. For this reason, we also model school effects by grade to avoid restricting the model parameters and to flexibly capture these SEL differences across grades.

Moreover, it is worth noting that equation (3) does not include a teacher-level effect. Therefore, the school effect α_{cjt}can be interpreted as an average effect of all teacher effects in school *j* and all other school-level factors, such as school culture and climate, school resources, and even neighborhood factors. The model specification is common in research (Deming, 2014; Meyer & Dokumaci, 2015; Chiang et al., 2016; Ehlert et al., 2016; Angrist et al., 2016, 2017) as well as for state school performance measurement systems (Education Analytics, 2017; SAS, 2018).

Equation (3) is estimated using the errors-in-variables (EIV) regression (Fuller, 1987) to to avoid producing biased estimates as a result of measurement error in the lagged academic and SEL measures. This approach corrects the sums-of-squares-and-cross-products matrix with an estimate of the variance of the measurement error in the right-hand-side variables. The correction transforms the sums-of-squares-and-cross-products matrix to reflect the expected variances and covariances of the right-hand-side variables in absence of measurement error. We use Cronbach's alpha for the lagged SEL scale scores and IRT conditional standard errors of measurement for lagged SBAC scores as measures of the variance of measurement error, respectively.[6]

The school effects are centered to have a weighted mean of zero within district and year for comparability. The weights correspond to the number of students in each school in the regression sample. Hence, the school effects are relative to the district and year average of school

[5] We compared school effects in 2016-17 calculated with this model to a model that includes two years of previous test scores. The correlations of school effects across these two models are close to one suggesting that adding an additional lag of prior scores does not substantially improve identification of school effects. See Figure A.2 in the appendix for the results.

[6] See Table A.1 in the appendix for the reliability values.

effects. This centering takes into account idiosyncratic shocks in a given district and year, such as differences in survey administration or district-wide (but not state-wide) events (e.g., a teacher's strike). Moreover, the current and lagged SEL scale scores and the SBAC scores were standardized to have a mean of zero and a standard deviation of one in each regression sample. Therefore, the coefficients on the lagged scores can be interpreted as a standard-deviation change in the outcome in response to a standard-deviation change of the lagged scores.

We estimate equation (3) for grades five through 12. Because SBAC is only administered in grades three through eight and again in grade 11, we can report results for math and ELA scores for grades five through eight. We do not report results for math and ELA in grade 11, because we cannot control for prior math and ELA scores for the SEL models in grade 10. This lack of prior tests raises the concern that we may omit other factors that influence the SEL constructs captured by test scores in year *t*-1. In that case, we could not compare estimates in grades five through nine and twelve to estimates in grades 10 and 11. To address this concern, we present the within-school R^2 for models using all available lagged outcomes, as well as models using only lagged SEL outcomes. The within-school R^2 is an adaptation of the R^2 that uses only the within-school component of the variances of the outcome, predictions, and residuals. It measures how much of the variance in the outcome is explained by covariates other than the school effect. Hence, the comparison of the R^2 of the two models for the same outcomes allows us gauge whether we are missing important factors by omitting SBAC scores.

We adjust the variance estimate of the school effects to correct for error in the SEL constructs. We estimate the noise-corrected variance of estimated school effects in the SEL growth models as follows:

$$Est.Variance\left[\alpha_{cjt}\right] = Sample\ Variance\left[\hat{\alpha}_{cjt}\right] - Sample\ Mean[\hat{\sigma}^2_{cjt}] \qquad (4)$$

where α_{cjt} are the effects for school *j* on outcome *c* at time *t*; $\hat{\alpha}_{cjt}$ are the estimated school effects from equation (3) and centered to have a mean of zero; and $\hat{\sigma}^2_{cjt}$ are the squared standard error estimates of the estimated and centered school effects. This approach estimates the variance of the true school effects α_{cjt} without variance due to the estimation error.

To measure stability of school effects, we report two correlations of school effects over time. First, we report correlations between the same grades in two different years. For example, we calculate the correlation between school effects in grade nine estimated for students in 2015-16 and the school effects in grade nine estimated for students in 2016-17. Second, we calculate the correlations of school effects in adjacent grades for a given cohort. These correlations provide evidence concerning whether schools have sustained impacts on students' SEL or whether perceived school effects reverse in the following year.

We show correlations both with and without noise correction to adjust for attenuation from sampling error in the school effect estimates. When computing the disattenuated correlations between school effects in the same grade across two years, we divide the unadjusted correlations by the square root of the product of the reliabilities of the school effects across the two years. We estimate the reliabilities of the school effects within each grade and year using the following equation:

$$Avg.Reliability\left[\hat{\alpha}_{cjt}\right] = Est.Variance\left[\alpha_{cjt}\right] / Sample\ Variance\left[\hat{\alpha}_{cjt}\right]. \qquad (5)$$

The reliability is an estimate of the proportion of the variance of the school effect estimates that is not the result of sampling error. For the correlations between school effects in adjacent grades in consecutive years measured using the same cohort of students, we cannot use this adjustment, because the residuals ε_{cijt} for individual students in that cohort are correlated from year to year. To adjust for measurement error in this case, we have to take into account correlation of the sampling errors between school effects. We do this in three steps. First, we compute adjusted variances in the school effects by grade and year using equation (4). Second, we compute an adjusted covariance using a covariance analogue of (4) that subtracts from the sample covariance the mean of the estimated covariances between the sampling errors across the two sets of school effects. Third, we compute the adjusted correlation by dividing the adjusted covariance by the square root of the product of the adjusted variances of the two sets of school effects. In practice, we do not have a perfect measure of measurement error. In both cases, the noise correction accounts for estimation error due to randomness in student-level SEL growth. School-level estimation error and error in the underlying measure not due to the internal consistency of the measures, however, are not accounted for.

Results

Model Coefficients and Goodness-of-Fit

Before assessing the stability of school effects on SEL, we begin by presenting the model coefficients and the goodness-of-fit of the school value-added models. Figure 2 displays the model coefficients of the lagged outcomes for the SBAC scores and the four SEL constructs separately for students in 2015-16 and 2016-17. Since all scores are standardized to have a mean of zero and a standard deviation of one, the effect sizes can be interpreted as a one standard-deviation increase in the outcome in response to a one-standard-deviation increase of the respective lagged score.

Across SEL constructs and grades, the coefficients of the same-outcome lagged variable are substantially larger than those of the other lagged constructs in the model (for example, the coefficient for prior growth mindset is larger than the coefficients for the other three SEL pretest measures when growth mindset is the outcome). For the four SEL constructs, these same-construct coefficients range from 0.29 to 0.55 for growth mindset, 0.36 to 0.54 for self-efficacy, 0.41 to 0.52 for self-management, and 0.45 to 0.57 for social awareness. For the SBAC scores, the coefficients of the same-subject lagged scores are bigger than for the SEL constructs, ranging from 0.71 to 0.77 for ELA and 0.71 to 0.98 for math. The coefficients of the other academic subject (i.e., prior ELA in the math outcome model and prior math in the ELA outcome model) are generally similar to the prior SEL measures. Furthermore, coefficients do not change notably from students in school years 2015-16 to students in 2016-17.

Figure 2: Model Coefficients

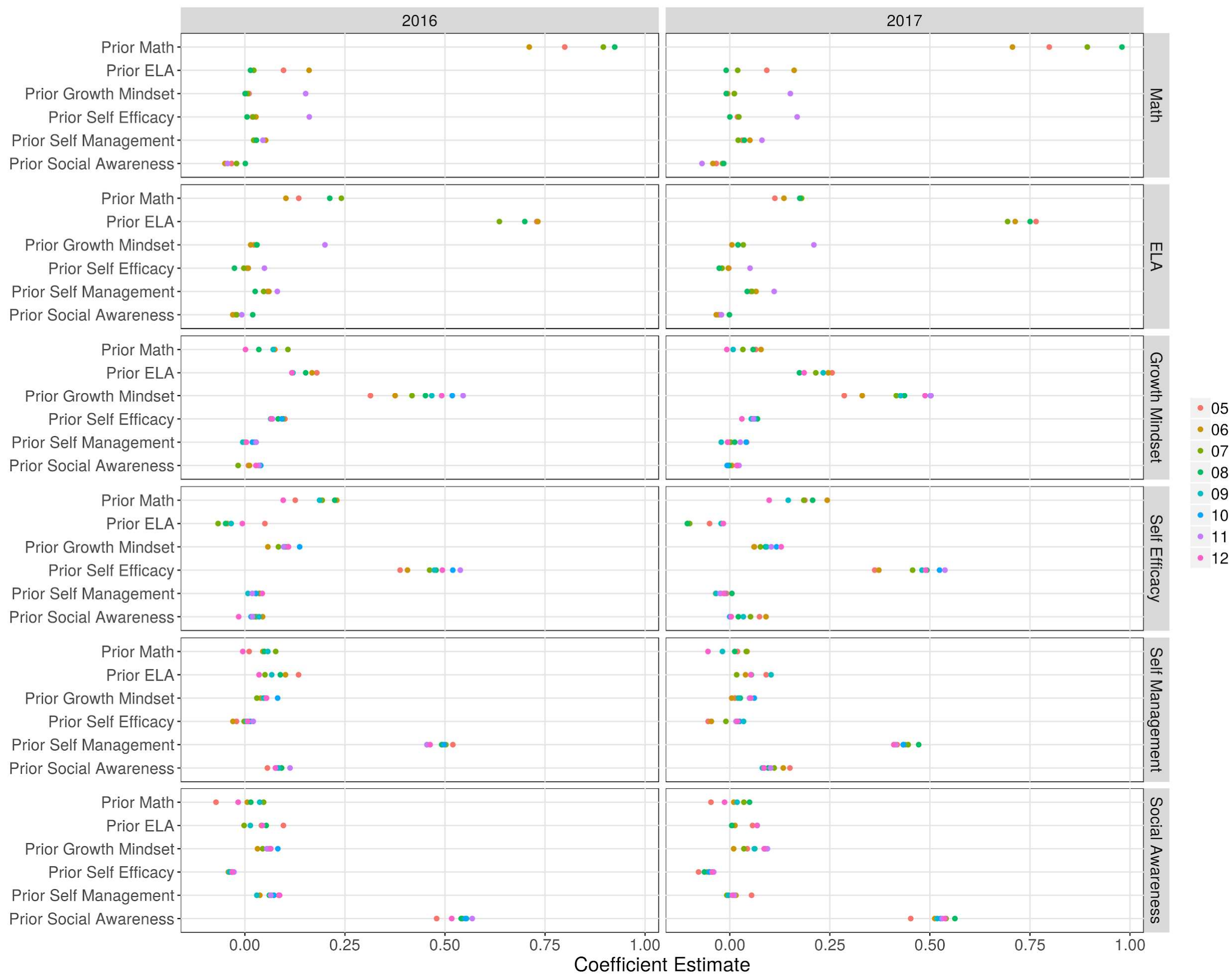
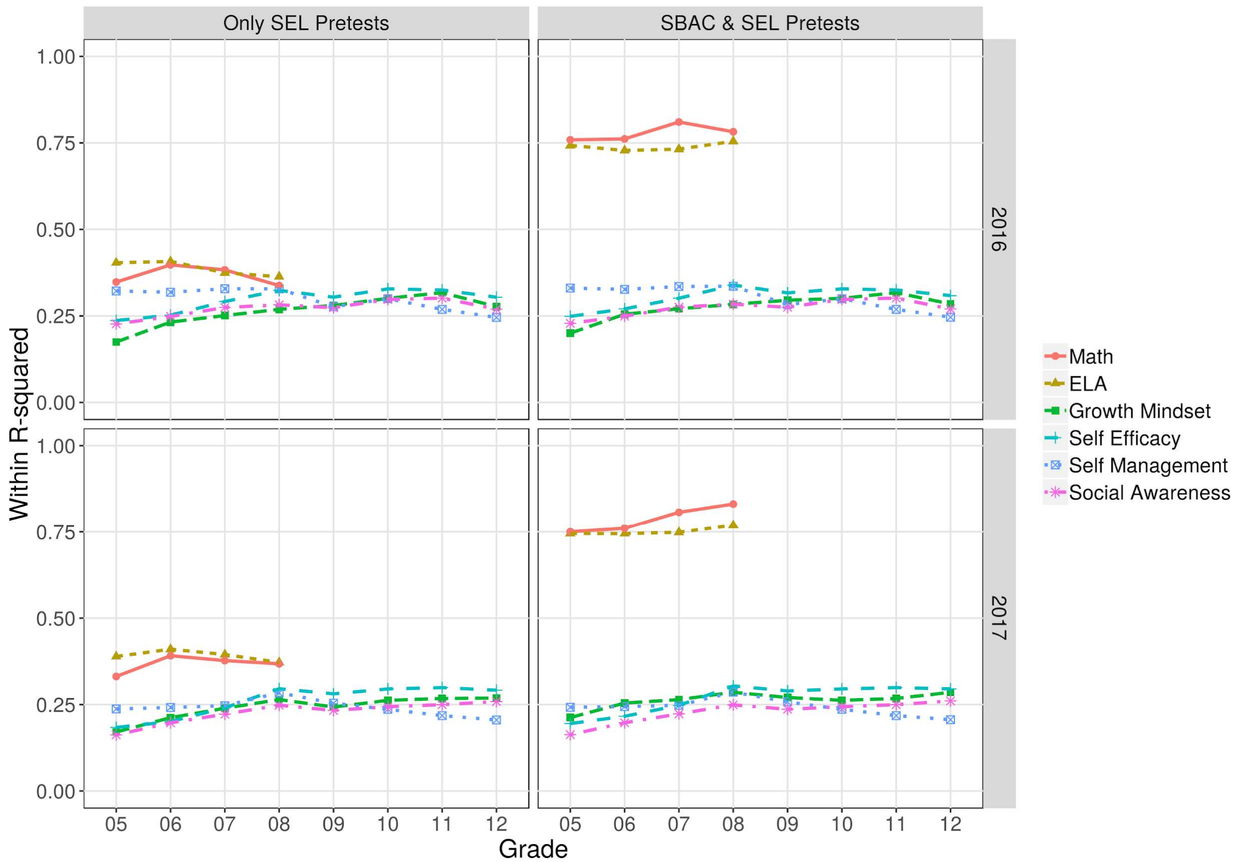


Figure 3 presents the within-school R^2 for each model. The within-school R^2 measures how much of the variation in the outcome is explained by covariates in the model beyond the school effects. The right panel includes all available pretest scores and demographic characteristics. In line with findings from Loeb et al. (2019), the within-school R^2 is substantially lower in models with SEL outcomes than in those with SBAC outcomes, raising the concern that other important factors may be omitted when modelling school effects on SEL. This pattern is present in both years.

Figure 3. Within-School R^2 by Grade



The within-school R^2 values allow us to assess whether the SEL models in grades 10 and 11 are comparable to those in other grades. We cannot include the same set of covariates for the SEL models in grades 10 and 11, because California does not administer the SBAC in grades nine and 10. The left panel of Figure 3 provides the within-school R^2 for every model only including lagged scores of the SEL constructs and demographic characteristics. Therefore, the R^2 measures for grades 10 and 11 are based on the same model specification and are numerically identical in both the left and the right panel. The R^2 measures for SEL constructs in the other grades are remarkably similar. The similarity suggests that SBAC scores do not explain much variation in SEL outcomes beyond lagged SEL scores and demographic characteristics, and that SEL models for grades 10 and 11 are comparable to models for other grades.

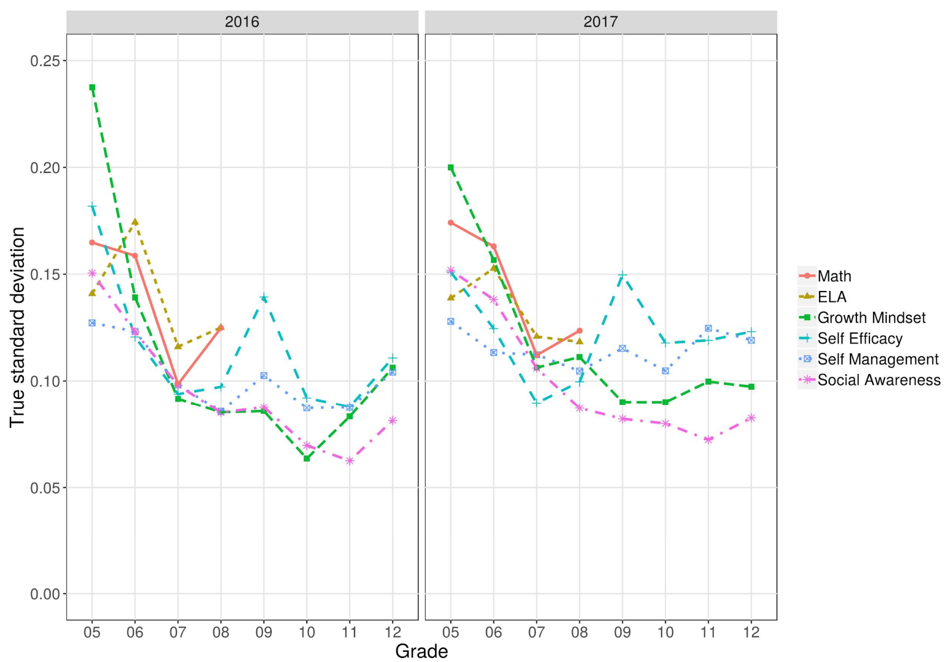
Variance of School Growth Estimates

Figure 4 displays the noise-corrected standard deviations for all outcomes across grades in years 2015-16 and 2016-17.[7] As in Loeb et al (2019), the variance in school effects on SEL constructs and SBAC scores are similar. Furthermore, the standard deviations of school effects

[7] See Table A.2 in the appendix for the corresponding standard deviations.

on SEL constructs follow a similar progression across grades for each construct. The standard deviations start out relatively high in grade five (0.24 and 0.20 for growth mindset, 0.18 and 0.15 for self-efficacy, 0.13 and 0.13 for self-management, and 0.15 for social awareness in years 2015-16 and 2016-17, respectively) and then decline until they level off or begin to slightly increase in high school (0.11 and 0.10 for growth mindset, 0.11 and 0.12 for self-efficacy, 0.10 and 0.12 for self-management, and 0.08 and 0.08 for social awareness in grade 12 in years 2015-16 and 2016-17, respectively). This decline might be explained by the increasing number of teachers who interact with students in higher grades. Although in elementary school just one high- or low value-added teacher may influence students' SEL, in middle and high school, contributions of different high and low value-added teachers may average out. Grade nine shows a notable spike in the standard deviation of school effects on self-efficacy in both years (0.14 and 0.15 in years 2015-16 and 2016-17 respectively). Generally, the patterns are remarkably consistent between years.

Figure 4. Noise-Corrected Standard Deviations of School Effects by Grade and Year



Correlations of School Growth Measures Across Years

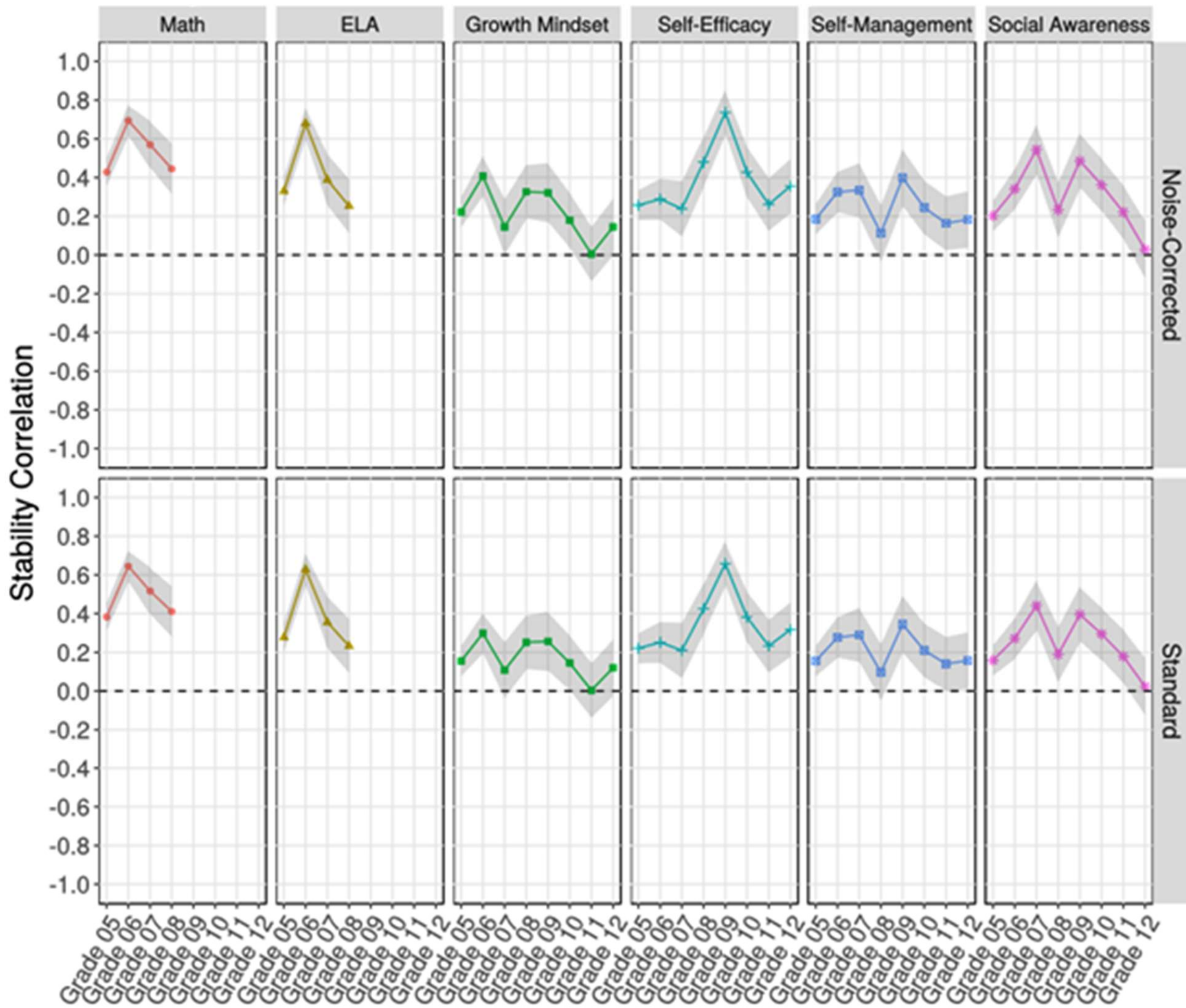
Next, we turn to the question of whether schools that we estimate to influence student SEL in one year also appear to influence student SEL in the next year. Figure 5 displays the standard and the noise-corrected correlations of school effects in the 2015-16 school year and school effects in the 2016-17 school year.[8] Gray shaded areas highlight 95% confidence intervals. The noise-corrected correlations are slightly larger than the standard ones, but overall are similar,

[8] See Table A.3 in the appendix for the corresponding correlations.

not surprisingly, as the reliability estimates are all close to one. We therefore refer to only the noise-corrected correlations from this point forward.

The correlations for the SBAC math and ELA scores increase from grade five to six and then decrease again in grades seven and eight; they range from 0.43 to 0.69 for math and from 0.25 to 0.68 for ELA. Estimates for growth mindset, self-management, and (slightly less so) self-efficacy follow the same pattern as correlations for school effects on math and ELA scores in grades five through seven. However, their magnitude is smaller (ranging from 0.14 to 0.41). In fact, correlations for all four SEL constructs decline from grade eight through 12 and are even indistinguishable from zero at the five percent significance level in a few grades. The correlations of school effects on self-efficacy and social awareness stand out. For self-efficacy, the correlations increase from grade seven (0.24) to grade nine (0.74) and then decrease again in grades 10 (0.43) and 11 (0.26); for social awareness, the correlations peak in grade seven (0.44) and grade nine (0.40). Overall, these correlations show that estimated school effects on self-reported SEL are mostly significantly positive but not particularly stable from one year to the next year.

Figure 5. Standard and Noise-Corrected Correlations Across Years

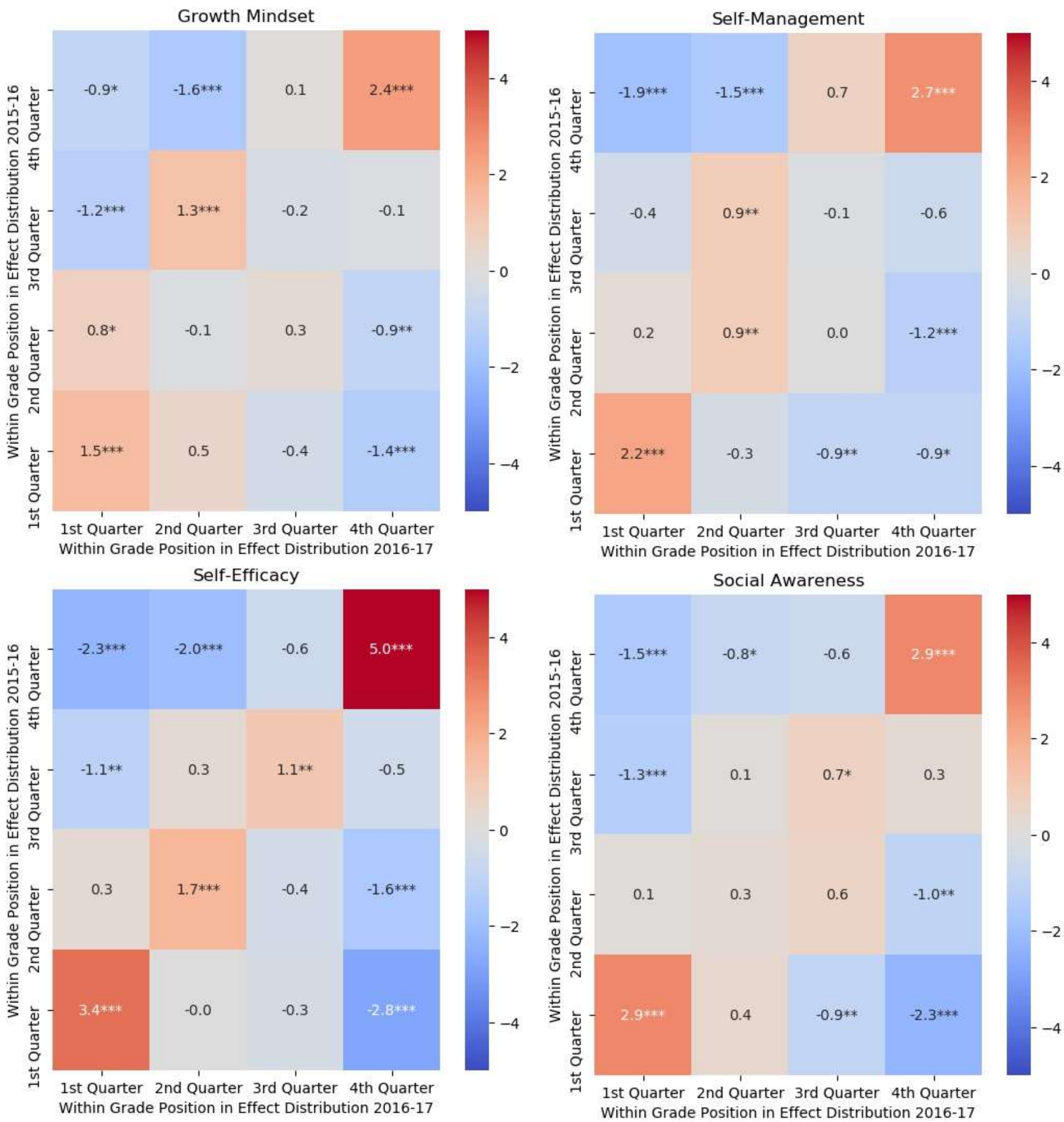


To understand whether certain schools consistently perform well or poorly, we examine the movement of schools along the school effect distribution from 2015-16 to 2016-17. Figure 6 shows transition matrices based on a school's position in the school effect distribution by quarters in 2015-16 and 2016-17 for all four SEL constructs.[9] The first quarter corresponds to the lowest quarter in the distribution, and the fourth quarter to the top quarter in the distribution. The numbers indicate the percentage point deviation from the percent of schools in each quadrant that we would observe in a totally random process (6.25%). We pool the graphs across grades but calculate the position in the distribution within grade. The significance levels derive from non-parametric permutation *p*-values. Overall, for all four SEL constructs, the percent of schools transitioning from the second or third quarter of the school effect distribution in 2015-16 to the same quarter in the distribution in 2016-17 is close to random. However, a significant

[9] Only school-by-grade effects with at least 20 students in the school and grade were used. Results are similar if all school effects are included.

greater percentage of schools are persistently in the top and bottom quarter in all four constructs than would be expected through random allocation. These results point to the possibility that while value-added measures of schools' contribution to survey-based SEL measures may not distinguish most schools from each other, they are able to distinguish a group of schools that are consistently either low-performing or high-performing in SEL growth.

Figure 6: Within-Grade Rank Transitions from 2015-16 to 2016-17



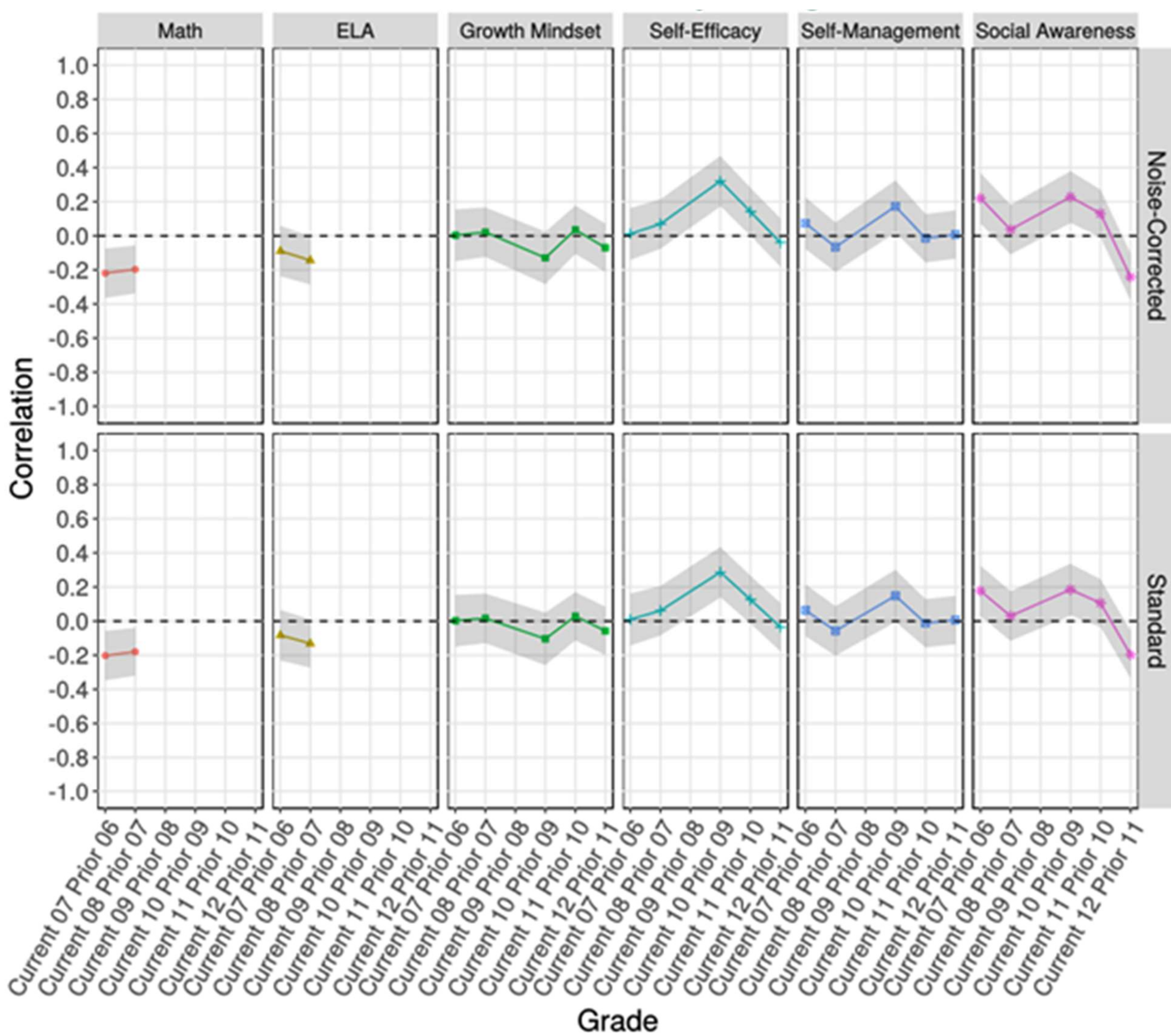
Correlations of School Growth Measures Grades within Cohort

Next, we estimate the extent to which schools identified to effectively support students' SEL in one cohort in one year do so again for the same cohort in the next year based on school value-added on self-reported SEL. To answer this question, we calculate the correlation of school effects in adjacent grades between years. For example, we correlate school effects for students in grade seven in school year 2015-16 with school effects on the same construct for students in grade eight in year 2016-17. We do not report the correlations for students moving from elementary school to middle school (grade five to six) and from middle school to high school (grade eight to nine) since students generally change schools during these transitions.

Figure 7 presents these correlations.[10] First, consider the correlations for SBAC math and ELA scores. For both measures, the correlations are small and negative, ranging from -0.22 to -0.09. The correlations for ELA are not significantly different from zero at the five percent significance level. The correlation of school effects on the four SEL constructs are mostly positive but statistically insignificant, with the exception of the correlation coefficients for (i) self-efficacy between grade ten and nine (0.32), (ii) self-management between grade ten and nine (0.17), and (iii) social-awareness between both grades seven and six (0.22) and ten and nine (0.23), which are significantly different from zero. Overall, these results suggest that, based on school value-added on SEL survey responses, a positive school effect in one year does not necessarily mean that the same students will experience a positive effect in the following year, though they are not more likely to experience a negative effect.

[10] See Table A.4 for the specific values of the correlations.

Figure 7. Cohort Correlations Between 2016 and 2017 Adjacent Grades



Conclusion

In this paper, we examine the stability of school effects on students' growth in SEL using a unique large-scale panel survey from the CORE districts. Their dataset contains responses to a survey eliciting four SEL constructs, growth mindset, self-efficacy, self-management, and social awareness, for more than 400,000 students each year across three years. We use a methodology similar to that of conventional value-added models of standardized assessments to assess variation across schools for each construct. We do this separately for two years (2015-16 and 2016-17) and then compare the results of these two models across years.

We find that the correlations among school effects on SEL across the two years are positive but generally low, and they are lower than those for math and ELA test scores. Moreover, the correlations of school effects in adjacent grades for the same cohort are mostly statistically insignificant. These results suggest that although the value-added measures may capture true school effects on SEL, the measures are also unstable from year to year.

This lack of stability could have two explanations. First, schools may not have the kinds of effects on their students' social-emotional development that they can reproduce over several years (i.e., a year or more). For instance, unlike math and ELA instruction, schools may not be explicitly teaching SEL; thus, they may not have established strategies and practices that are consistently implemented from one year to the next. However, this explanation stands in contrast to the accumulating evidence that schools and school-based interventions can have important effects on student SEL (Blum, et al., 2004; Battistich et al., 2004; Hamre & Pianta, 2006; Allensworth & Easton, 2007; Jennings & Greenberg, 2009; Durlak, et al., 2011; McCormick et al., 2015; Berkowitz et al., 2016). Second, the share of the variance in the estimated school effects actually explained by school practices may be lower than the models suggest. The estimated effects of schools may result not only from school effects but also unaccounted measurement error. Sampling error and the internal consistency-based error from the underlying measure are not the only possible sources of measurement error. Students may consistently report their assessment of their SEL on the survey but this assessment might be error prone, even varying from day to day. Lagged outcomes and demographic characteristics explain less of the within-school variation of SEL scales than for SBAC math and ELA test scores, which may reflect noise in the outcome.

In either case, our results suggest that school value-added measures on students' self-reported SEL exhibit volatility across years; this volatility is higher than for growth on academic assessments. This finding highlights potential drawbacks of using such self-report measures of SEL for school accountability purposes with sanctions or rewards attached to them. Because of the low stability, schools would likely be punished or rewarded for factors that are outside of their control (Kane and Staiger, 2002a). The results in this paper also suggest that even the identification of best school-wide practices or the allocation of resources and supports tied to measures of school value-added on SEL on a yearly basis may be problematic. Because schools identified as effective in supporting SEL in one year may not be identified as such in the next year, educational policies and practices attributed to the success of schools may hence change at a similar rate (Kane & Staiger, 2002a). Similarly, administrators may have to reallocate support and resources abruptly from one year to the next.

To date, school performance measurement systems do not include school value-added models of SEL. Instead, the CORE districts report school-by-grade levels overall and by subgroup for the four SEL constructs. Although SEL levels have lower stability over time than do SBAC scores, they are more stable than SEL school value-added measures (White and Polikoff, 2019). The greater stability of levels compared to gains may appear to indicate that levels of SEL are more reliable measures of schools' contributions to SEL. However, levels do not control for student characteristics, including students incoming SEL levels, and thus, are unlikely to accurately identify schools that effectively (or ineffectively) support the social-emotional development of their students.

Our findings indicate that value-added measures, though conceptually superior to levels, are not stable enough to distinguish the contribution of most schools to their students' social-emotional development. We do find, however, that some schools rank persistently in the top or bottom quarter of the distribution of school effects in both years. These thicker than expected tails of the distribution provide some evidence that, although value-added on SEL does not appear to distinguish well among most schools, a group of schools does stand out. This finding is in line with the results for school's value-added to test scores. Kane and Staiger (2002b) show that school value-added has substantial measurement error, leading to an inability to consistently distinguish most schools. However, other research has shown that some schools do stand out as either particularly good or particularly bad on measures of value added to academic achievement gains (Angrist, Cohodes, Dynarski, Pathak & Walters, 2016; Abdulkadiroğlu, Angrist, Hull & Pathak, 2016).

The results reported here are preliminary, in that there are a host of other research questions we aim to pursue to better understand the stability or volatility of these measures over time. For instance, examining additional or alternative approaches for measurement error correction in models using students' SEL, such as finite sample approaches, may be important for reducing bias in these models. Our recent efforts to apply a similar approach described in this paper to classroom-level impacts on students' SEL suggest there may be more variance explained at the classroom level than the school level, even after accounting for the school; additional research is needed to examine whether such classroom effects are more stable than school-wide effects. Finally, it is worth noting that these conclusions are based on three years of data (i.e., two years of growth). Although the results presented here highlight important findings, including more years of data may deepen our understanding of the stability of these measures over time.

References

Allensworth, E. M., & Easton, J. Q. (2007). *What matters for staying on-track and graduating in Chicago public high schools: A close look at course grades, failures, and attendance in the freshman year*. Chicago, IL: Consortium on Chicago School Research.

Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. D. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 4) (pp. 1-181). Amsterdam: Elsevier, North-Holland.

Abdulkadiroğlu, A., Angrist, J., Hull, P., & Pathak, P. (2016). Charters without Lotteries: Testing Takeovers in New Orleans and Boston, *American Economic Review*, 106(7), 1878-1920.

Angrist, J., Cohodes, P., Dynarski, S., Pathak, P., & Walters, C. (2016). Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice, *Journal of Labor Economics*, 34(2), 275-318

Angrist, J., Hull, P., Pathak, P., & Walters, C. (2016). Interpreting tests of school VAM validity. *American Economic Review: Papers and Proceedings*, 106(5), 388-392.

Angrist, J., Hull, P., Pathak, P., & Walters, C. (2017). Leveraging lotteries for school value-added: Testing and estimation. *Quarterly Journal of Economics*, 132(2), 871-919.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W.H. Freeman.

Battistich, V., Schaps, E., & Wilson, N. (2004). Effects of an elementary school intervention on students' “connectedness” to school and social adjustment during middle school. *Journal of Primary Prevention*, 24(3), 243–262.

Berkowitz, R., Moore, H., Astor, R. A., & Benbenishty, R. (2016). A research synthesis of the associations between socioeconomic background, inequality, school climate, and academic achievement. *Review of Educational Research*, 87(2), 425-469.

Blum, R. W., Libbey, H. P., Bishop, J. H., & Bishop, M. (2004). School connectedness–strengthening health and education outcomes for teenagers. *Journal of School Health*, 74(7), 231–235.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students’ attitudes and behaviors. Educational evaluation and policy analysis, 39(1), 146-170.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633–2679.

Chiang, H., Lipscomb, S., & Gill, B. (2016). Is school value added indicative of principal quality*?* *Education Finance and Policy*, 11(3), 283-309.

Claro, S. & Loeb, S. (2019a). “Effect of Growth Mindset on Achievement Evidence from California CORE school districts.”, MIMEO

Claro, S. & Loeb, S. (2019b). “Students with Greater Self-Management Skills Learn More In School: Evidence from California CORE school districts”, MIMEO

Collaborative for Academic, Social, and Emotional Learning [CASEL]. (2005). *Safe and sound: An educational leader’s guide to evidence-based social and emotional learning programs—Illinois edition*. Chicago, IL.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738-782.

Data Quality Campaign. (2019). *Growth data: It matters, and it's complicated*. Washington, DC: Data Quality Campaign.

Dee, T. S., & West, M. R. (2011). The non-cognitive returns to class size. *Education Evaluation and Policy Analysis*, 33(1), 23-46.

Deming, D. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, 132(4), 1593-1640.

Deming, D. (2014). Using school choice lotteries to test measures of school effectiveness. *American Economic Review: Papers and Proceedings*, 104(5), 406-411.

Dumay, X., Coe, R., & Anumendem, D.N. (2014). Stability over time of different methods of estimating school performance, *School Effectiveness and School Improvement*, 25(1), 64-82.

Durlak, J. A., Dymnicki, A. B., Taylor, R., Weissberg, R. P., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development* 82(1), 405-432.

Dweck, C. (2006). *Mindset: The new psychology of success*. New York, NY: Random House Publishing Group.

Education Analytics (2017), *Growth Model for Educator Evaluation 2016/17 Technical Report*, prepared for the New York State Education Department by Education Analytics, Inc.

Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2016). Selecting growth measures for use in school evaluation systems: Should proportionality matter? *Educational Policy*, 30(3), 465-500.

Fuller, W. (1987). *Measurement error models*. Hoboken, N.J.: John Wiley & Sons.

Gehlbach, H., & Hough, H., (2018). Measuring social emotional learning through student surveys in the CORE Districts: A pragmatic approach to validity and reliability. *Policy Analysis for California Education*, Stanford, CA.

Gershenson, S. (2016). Linking teacher quality, student attendance, and student achievement. *Education Finance and Policy*, 11(2), 125-149.

Gray, J., Goldstein, H., & Thomas, S. (2001). Predicting the future: The role of past performance in determining trends in institutional effectiveness at A-level. *British Educational Research Journal*, 27, 1–15.

Grissom, J., Kalogrides, D., & Loeb, S. (2015). Using student test scores to measure principal performance. *Educational Evaluation and Policy Analysis*, 37(1), 3-28.

Guarino, C., Reckase, M., & Wooldridge, J. (2015). Can value-added measures of teacher performance be trusted? *Education Finance and Policy*, 10(1), 117-156.

Hamre, B. K., & Pianta, R. C. (2006). Student–teacher relationships. In G. G. Bear & K. M. Minke (Eds.), *Children's needs III: Development, prevention, and intervention* (pp. 59-71). Washington, DC: National Association of School Psychologists.

Heckman, J. J., & Kautz, T. (2013). Fostering and measuring skills: Interventions that improve character and cognition. In J. J. Heckman, J. E. Humphries, & T. Kautz (Eds.). *The myth of achievement tests: The GED and the role of character in American life* (pp. 341-430). Chicago: University of Chicago Press.

Heckman, J. J, Humphries, J. E., & Kautz, T. (Eds). (2014). *The myth of achievement tests: The GED and the role of character in American life*. Chicago: University of Chicago Press.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. American Economic Review, 91(2), 145-149.

Jackson, K. (2018). What do test scores miss? The importance of teacher effects on non-test-score outcomes. *Journal of Political Economy, 126*(5), 2072-2107.

Jennings, J. L., & DiPrete, T. A. (2010). Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education*, 83(2), 135–159.

Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, 79(1), 491–525.

Kane, T., & Staiger, D. (2002a). Volatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), *Brookings Papers on Education Policy* (pp. 235-283). Washington, D.C.: Brookings Institution.

Kane, T., & Staiger, D. (2002b). "The Promise and Pitfalls of Using Imprecise School Accountability Measures." *Journal of Economic Perspectives*, 16(4), 91-114.

Koedel, C., & Betts, J. (2010). Value added to what? How a ceiling in the testing instrument influences value-added estimation. *Education Finance and Policy*, 5(1), 54-81.

Koedel, C., Mihaly K., & Rockoff, E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47, 180-195.

Ladd, H. F., & Sorensen, L. C. (2017). Returns to teacher experience: Student achievement and motivation in middle school. *Education Finance and Policy*, 12(2), 241-279.

Liu, J. & Loeb, S. (2018). Engaging teachers: Measuring the impact of teachers on student attendance in secondary school (CEPA Working Paper No. 17-01). Retrieved from http://cepa.stanford.edu/wp17-01

Loeb, S., Christian, M. S., Hough, H., Meyer, R. H., Rice, A. B., & West, M. R. (2019). School differences in social-emotional learning gains: Findings from the first large-scale panel survey of students. *Journal of Educational and Behavioral Statistics.* https://doi.org/10.3102/1076998619845162

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Marks, G.N. (2015). The size, stability, and consistency of school effects: evidence from Victoria, *School Effectiveness and School Improvement*, 26(3), 397-414

McCaffrey, D., Lockwood, J.R., Koretz, D., Louis, T., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1), 67-101.

McCormick, M. P., Cappella, E., O’Connor, E. E., & McClowry, S. G. (2015). Context matters for social emotional learning: Examining variation in program impact by dimensions of school climate. American Journal of Community Psychology, 56(1-2), 101–119.

Meyer, R. (1997). Value-added indicators of school performance: A primer. *Economics of Education Review*, 16(3), 283-301.

Meyer, R. & Dokumaci, E. (2015). Value-Added Models and the next generation of assessments. In R. W. Lissitz (Ed.), *Value added modeling and growth modeling with particular* *application to teacher and school effectiveness*, Charlotte, NC: Information Age Publishing, 2015.

Meyer, R. H., Wang, Y. C., & Rice, A. B. (2018). Measuring students' social-emotional learning among California's CORE districts: An IRT modeling approach. *Policy Analysis for California Education*, Stanford, CA.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM algorithm. (ETS Research Report 92-06). Princeton, NJ: Educational Testing Service.

Raudenbush, S., & Willms, D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307-355.

Reardon, S., & Raudenbush, R. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4(4), 492-519.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.

Ruzek, E. A., Domina, T., Conley, A. M., Duncan, G. J., & Karabenick, S. A. (2015). Using value-added models to measure teacher effects on students' motivation and achievement. *The Journal of Early Adolescence*, 35(5-6), 852–882.

SAS (2018). *SAS EVAAS: Statistical Models and Business Rules of OH EVAAS Analyses (Prepared for Ohio Department of Education).* Retrieved from http://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Ohio-Report-Cards/Value-Added-Technical-Reports-1/Technical-Documentation-of-EVAAS-Analysis.pdf.aspx

Teddlie, C., & Reynolds, D. (2000). The international handbook of school effectiveness research. London, UK: Falmer Press.

Tekwe, C., Carter, R., Ma, C.X., Algina, J., Lucas, M., Roth, J., Ariet, M., Fisher, T., & Resnick, M. (2004). An empirical comparison of statistical models for value-added assessment of school performance. *Journal of Educational and Behavioral Statistics*, 29(1), 11-36.

Thomas, S., Peng, W. J., & Gray, J. (2007). Modelling patterns of improvement over time: Value-added trends in English secondary school performance across ten cohorts. Oxford Review of Education, 33, 261–295.

West, M. R., Buckley, K., Krachman, S. B., & Bookman, N. (2018). Development and implementation of student social-emotional surveys in the CORE Districts. *Journal of Applied Developmental Psychology*, 55, 119-129.

West, M.R., Pier, L., Fricke, H., Hough, H., Loeb, S., Meyer, R. H., & Rice, A. (2018). *Trends in Student Social-Emotional Learning: Evidence from the CORE Districts*. Stanford, CA: Policy Analysis for California Education.

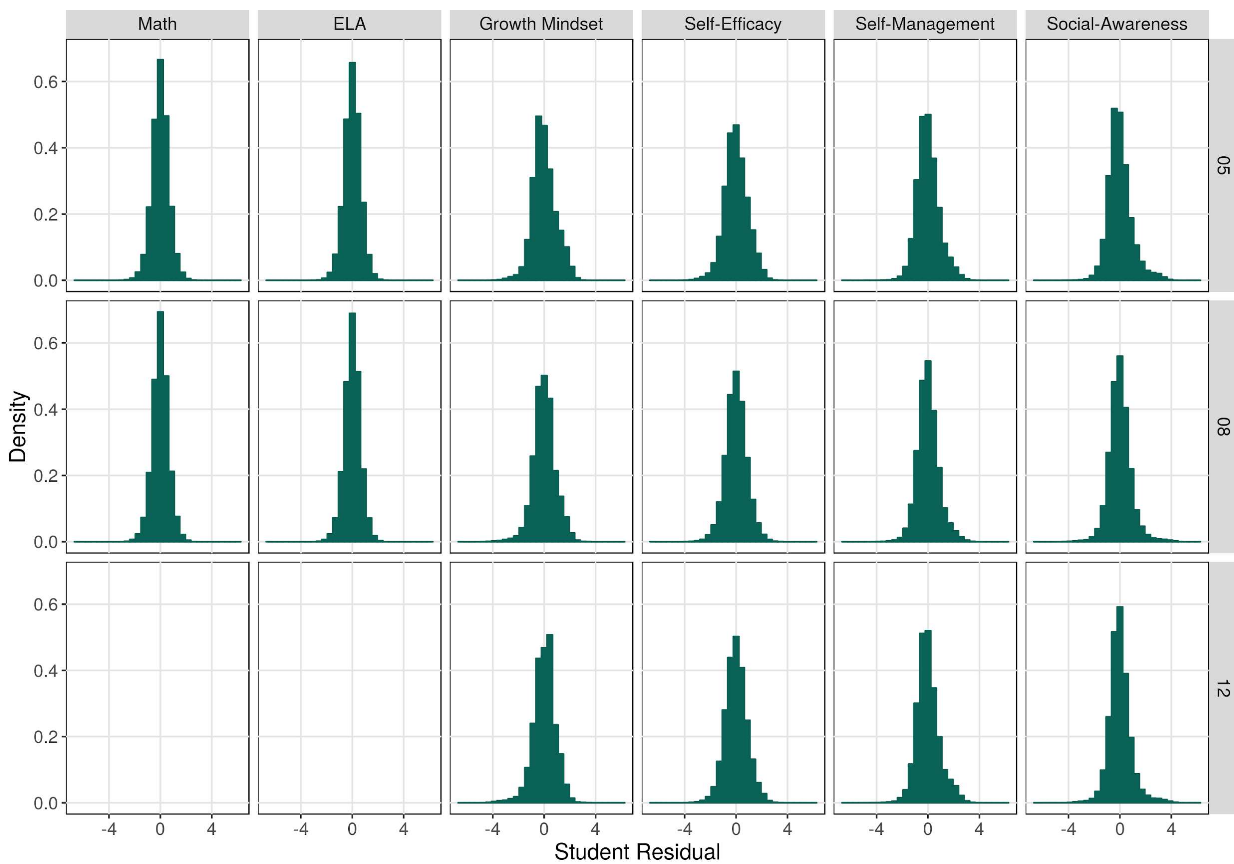
White, R. & Polikoff, M. (2019). "The Properties of Non-Academic School Performance Measures", Policy Analysis for California Education (PACE) Working Paper.

Appendix

Table A.1: Internal scale reliability of SEL and SBAC scale scores by grade

	4	5	6	7	8	9	10	11	12
Math	0.94	0.92	0.93	0.91	0.91				
ELA	0.92	0.92	0.91	0.92	0.91				
Growth Mindset	0.66	0.7	0.73	0.75	0.77	0.8	0.81	0.82	0.83
Self-Efficacy	0.83	0.86	0.87	0.88	0.89	0.89	0.89	0.89	0.89
Self-Management	0.82	0.84	0.85	0.87	0.87	0.86	0.86	0.86	0.86
Social Awareness	0.77	0.79	0.8	0.81	0.81	0.81	0.81	0.81	0.82

Figure A.1: Distribution of Student-Level Growth Scores for Grades 5, 8, and 12



Notes: We construct a student-level estimate of growth by computing the residual based on a prediction that excludes school effect α_{cjt} in Equation (3). This residual contains α_{cjt} plus the random student error ε_{cijt}.

Figure A.2: Correlations Between School Effects with One Year and Two Years

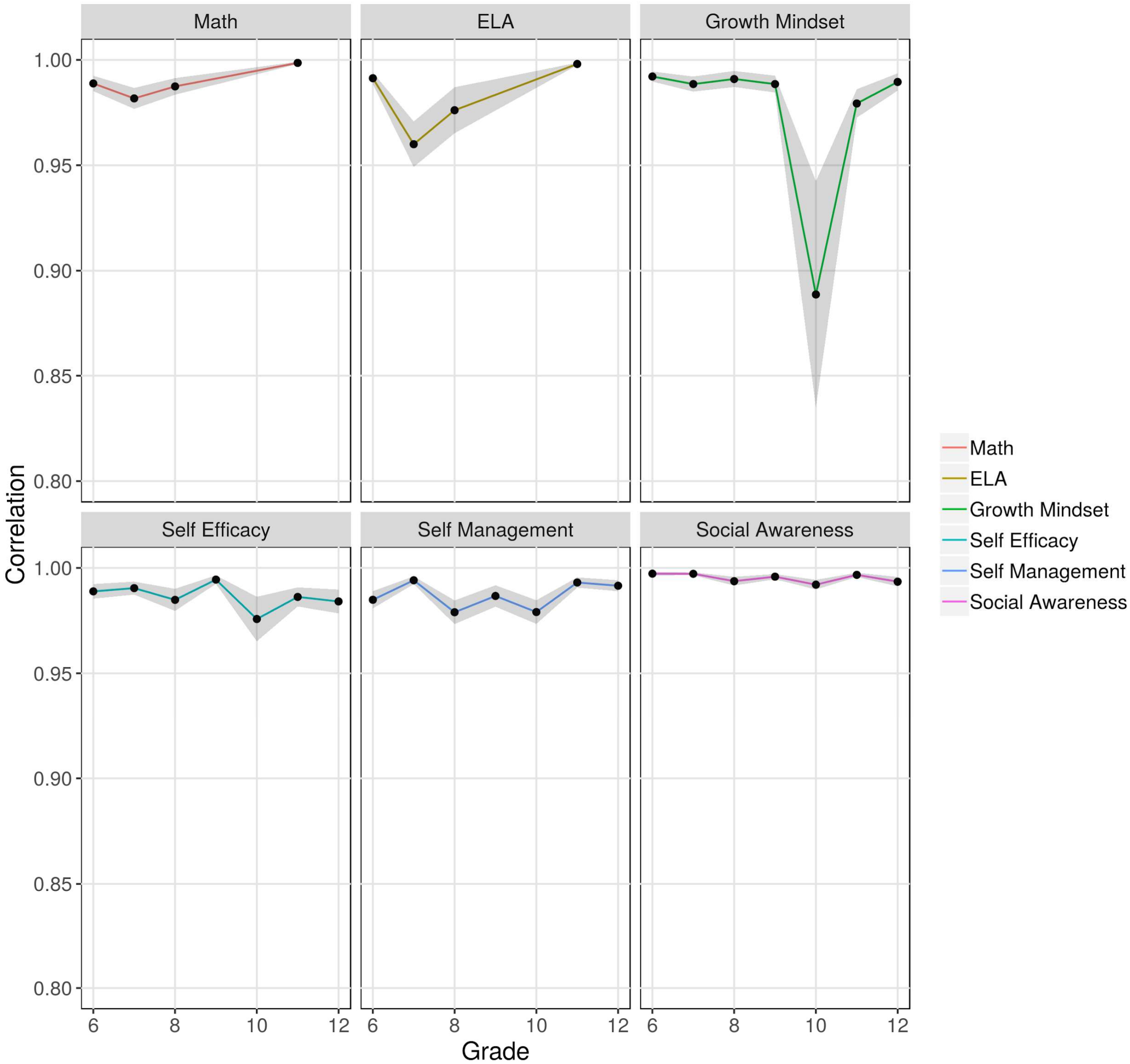


Table A.2: Standard Deviations of Noise-Corrected School Effects by Construct, Grade, and Year

Grade	Math	ELA	Growth Mindset	Self-Efficacy	Self-Management	Social Awareness
			2015-16			
5	0.14	0.16	0.24	0.18	0.13	0.15
6	0.17	0.16	0.14	0.12	0.12	0.12
7	0.12	0.10	0.09	0.09	0.10	0.10
8	0.13	0.12	0.09	0.10	0.09	0.09
9			0.09	0.14	0.10	0.09
10			0.06	0.09	0.09	0.07
11			0.08	0.09	0.09	0.06
12			0.11	0.11	0.10	0.08
			2016-17			
5	0.17	0.14	0.20	0.15	0.13	0.15
6	0.16	0.15	0.16	0.12	0.11	0.14
7	0.11	0.12	0.11	0.09	0.11	0.11
8	0.12	0.12	0.11	0.10	0.10	0.09
9			0.09	0.15	0.12	0.08
10			0.09	0.12	0.10	0.08
11			0.10	0.12	0.12	0.07
12			0.10	0.12	0.12	0.08

Table A.3 Correlations of School Effects in Same grade and Different Year

Standard						
Grade	Math	ELA	Growth Mindset	Self-Efficacy	Self-Management	Social Awareness
5	0.38	0.28	0.16	0.22	0.16	0.16
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
6	0.64	0.63	0.30	0.25	0.28	0.27
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
7	0.52	0.36	0.11	0.21	0.29	0.44
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
8	0.41	0.23	0.25	0.43	0.10	0.19
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
9			0.26	0.66	0.34	0.40
			(0.08)	(0.06)	(0.08)	(0.07)
10			0.15	0.38	0.21	0.29
			(0.07)	(0.07)	(0.07)	(0.07)
11			0.00	0.23	0.14	0.18
			(0.07)	(0.07)	(0.07)	(0.07)
12			0.12	0.32	0.16	0.02

Noise-Corrected						
Grade	Math	ELA	Growth Mindset	Self-Efficacy	Self-Management	Social Awareness
5	0.43	0.33	0.22	0.26	0.19	0.20
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
6	0.69	0.68	0.41	0.29	0.32	0.34
	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
7	0.57	0.39	0.14	0.24	0.33	0.54
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
8	0.45	0.25	0.33	0.48	0.11	0.23
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
9			0.32	0.74	0.40	0.49
			(0.08)	(0.05)	(0.07)	(0.07)
10			0.18	0.43	0.24	0.36
			(0.07)	(0.07)	(0.07)	(0.07)
11			0.00	0.26	0.16	0.22
			(0.07)	(0.07)	(0.07)	(0.07)
12			0.14	0.36	0.18	0.03
			(0.07)	(0.07)	(0.07)	(0.08)

Notes: Standard errors are in parentheses.

Table A.4 Correlations of School Effects in Adjacent Grades and Same Cohort

Standard						
Grade	Math	ELA	Growth Mindset	Self-Efficacy	Self-Management	Social Awareness
7 and 6	-0.20	-0.08	0.00	0.01	0.06	0.18
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
8 and 7	-0.18	-0.13	0.02	0.06	-0.06	0.03
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
10 and 9			-0.10	0.29	0.15	0.18
			(0.08)	(0.08)	(0.08)	(0.08)
11 and 12			0.03	0.13	-0.01	0.11
			(0.07)	(0.07)	(0.07)	(0.07)
12 and 11			-0.06	-0.04	0.01	-0.20
			(0.07)	(0.07)	(0.07)	(0.07)
Noise-Corrected						
Grade	Math	ELA	Growth Mindset	Self-Efficacy	Self-Management	Social Awareness
7 and 6	-0.22	-0.09	0.00	0.01	0.07	0.22
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
8 and 7	-0.20	-0.14	0.02	0.07	-0.07	0.04
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
10 and 9			-0.13	0.32	0.17	0.23
			(0.08)	(0.07)	(0.08)	(0.08)
11 and 12			0.04	0.14	-0.01	0.13
			(0.07)	(0.07)	(0.07)	(0.07)
12 and 11			-0.07	-0.04	0.01	-0.24
			(0.07)	(0.07)	(0.07)	(0.07)

Notes: Standard errors are in parentheses.