PACE
Policy Analysis for California Education
CORE-PACE RESEARCH PARTNERSHIP

WORKING PAPER

School Effects on Social-Emotional Learning: Findings from the First Large-Scale Panel Survey of Students

Measures of school-level growth in student outcomes are common tools used to assess the impacts of schools. The vast majority of these measures are based on standardized tests, even though emerging evidence demonstrates the importance of social-emotional skills (SEL). This paper uses the first large-scale panel surveys of students on SEL to produce and evaluate school-level value-added measures by grade for growth mindset, self-efficacy, self-management, and social awareness. We find substantive differences across schools in SEL growth, of magnitudes similar to those for academic achievement. This result suggests that schools might contribute to students' SEL. However, we also find that the models are not as well specified for SEL as they are for achievement gains, raising the possibility that the estimated school effects include school-level measurement error and potential omitted variables bias. In addition, the across-school variance in the average level of the SEL measures is proportionally much smaller than for academic measures, which would not be expected if substantial impacts of schools on SEL outcomes persisted over time. These findings recommend caution in interpreting measures as the causal impacts of schools on SEL, though they also do not rule out important school effects.

Susanna Loeb
Stanford University

Michael S. Christian
Education Analytics

Heather J. Hough
Policy Analysis for California Education

Robert H. Meyer
Education Analytics

Andrew B. Rice
Education Analytics

Martin R. West
Harvard University

VERSION: May 2018

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 the leaders and administrators in the CORE districts for their support throughout this project, and the generous funder of this research, the S. D. Bechtel, Jr. Foundation. PACE working papers are circulated for discussion and comment purposes and have not undergone the peer-review process that accompanies official PACE publications.

Introduction

State departments of education, as well as many school districts, use growth measures as tools to assess the impacts of schools and teachers. By one account, growth is used to measure school performance in forty-two of the fifty U.S. states and in the District of Columbia (Thomsen, 2013). Much of the literature on growth measures, which are also called value-added or academic progress measures, has focused on the teacher level (e.g., Rivkin, Hanushek & Kain, 2005; Chetty, Friedman & Rockoff, 2014). However, there have also been a substantial number of studies focused on school-level value-added. Topics covered by these studies include the conceptualization and estimation of school effects (Raudenbush & Willms, 1995; Meyer, 1997; Tekwe et al., 2004; Reardon & Raudenbush, 2009; Ehlert, Koedel, Parsons, & Podgursky, 2016); the implications of school growth measures in accountability systems (Ladd & Walsh, 2002; Kane & Staiger, 2002); the persistence of school value-added effects over time (Briggs & Weeks, 2011); the usage and adaptation of school value-added measures as tools to evaluate the impacts of principals (Grissom, Kalogrides, & Loeb, 2015; Chiang, Lipscomb, & Gill, 2016); and tests of the validity of school effects using data from school choice lotteries (Deming, 2014; Angrist, Hull, Pathak, & Walters, 2016, 2017).

The vast majority of studies of growth in education have modeled outcomes in academic subjects, such as in mathematics and English language arts, based on student performance on standardized tests. This focus on academic subjects exists in light of a substantial body of emerging research finding that social-emotional skills (sometimes called non-cognitive skills) contribute to school success and adult outcomes (Heckman & Rubenstein, 2001; Kautz, Heckman, Diris, ter Weel, & Borghans, 2014). Cohen, Garcia, Apfel, and Master (2006), for example, found that a brief in-class writing assignment affirming sense of personal adequacy significantly improved the grades of African American students and reduced the racial achievement gap. Similarly, Blackwell, Trzesniewski, and Dweck (2007) found that an intervention teaching an incremental theory of intelligence (“growth mindset”) to seventh graders increased reported classroom motivation and grades. These and other studies demonstrate that school performance depends on more than the knowledge and skills typically measured by standardized tests.

While many factors likely contribute to students’ social-emotional skills, research increasingly provides evidence that experiences in schools can affect social-emotional learning both directly (Allensworth & Easton, 2007; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011) and through the implementation of policies and practices that improve a school’s culture and climate and promote positive relationships (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). A meta-analysis by Durlak et al. (2011) finds that school programs and interventions, such as the ones studied by Cohen et al. (2006) and Blackwell et al. (2007), can improve social-emotional skills. While Kautz et al. (2014) note that the short follow-up of most studies of elementary school programs makes it difficult to draw strong conclusions about their long-term impacts (particularly in comparison to early childhood programs), they also note that evidence from the studies that have followed participants into adulthood is promising.

Recent studies also have shown that teachers can affect student social-emotional development (e.g., Jennings & DiPrete, 2010; Gershenson, 2016; Ladd & Sorensen, 2017; Liu & Loeb, 2018). For example, Ruzek, Domina, Conly, Duncan, & Karabenick (2015) demonstrate teacher effects on academic motivation, while Blazar and Kraft (2017) find effects on self-reported self-efficacy and happiness, and Jackson (in press) finds teacher effects on a composite measure of student GPA, on-time grade completion, suspensions, and full-day attendance. Teachers who increase test performance are not necessarily the same as those who help students improve their social-emotional skills. In fact, the correlations between teachers' effects on test scores and teachers' effects on non-test scores are weak (Jackson, in press; Liu & Loeb, 2018). More generally, a large portion of teacher effects on student long-term outcomes, like college attendance, is not explained by teacher effects on student achievement, suggesting that good teachers not only increase students' test scores, but also impact other outcomes (Chamberlain, 2013).

While some research has assessed teacher effects, none to date, of which we are aware, have assessed the extent to which schools at large vary in their students' social-emotional learning trajectories. Yet school-level differences, beyond differences across teachers, could impact student development of these skills. School leaders have been shown to affect student learning through mechanisms such as building a sense of community that could also affect students' social-emotional development (see Waters, Marzano, & McNulty, 2003; Leithwood, Louis, Anderson, & Wahlstrom, 2004; and Hallinger, 2005, for meta-analyses). Moreover, the prevalence of bullying and other culture and climate characteristics of schools can affect students and their social-emotional health and development (Olweus, 1994), and school-based interventions can affect these cultural characteristics (Ttofi & Farrington, 2011).

In this paper, using the first large-scale panel surveys of students on social-emotional learning (SEL) outcomes, we produce and evaluate school-level value-added measures by grade for four dimensions of SEL learning: growth mindset, self-efficacy, self-management, and social awareness. We measure student-level SEL outcomes using student responses to SEL-related items in 2015 and 2016 administrations of surveys, which we scale using Item Response Theory (IRT) methods. Given that these are the first two administrations of such a survey at a large scale, this is the first opportunity to measure differences in student growth in SEL outcomes across a variety of schools. The value-added measures cover the growth of more than 150,000 students in grades four through eight across schools in the CORE districts, large urban districts that began measuring SEL as part of a multiple-measures accountability system under a No Child Left Behind flexibility request.[1] For comparison, we also measure school value-added

[1] Fresno, Long Beach, Los Angeles, Oakland, San Francisco, and Santa Ana unified school districts were part of the NCLB waiver, and Garden Grove and Sacramento City unified school districts are also part of the CORE network.

measures in academic assessment scores in mathematics and English language arts (ELA) for the same students.

We assess these school-level growth models by measuring the reliability of the underlying variables, the fit of the models to the data, the variance in growth across schools, and the consistency of measures across grades within the same school, and by comparing these measures to similar ones produced for academic test performance. On the promising side, we find substantive differences across schools in growth in student SEL outcomes. Over the four SEL constructs and five grades, the estimated standard deviation of impacts across schools relative to the standard deviation of SEL outcome measures across students is between 0.09 and 0.24. This magnitude is similar to the estimated variance of school effects on academic achievement in math and ELA, which are between 0.11 and 0.18 across grades. This result suggests that schools might contribute to students' development of social-emotional skills. However, we also find that the value-added models are not as well specified for SEL as they are for achievement gains, with the covariates explaining far less of the variance across students in SEL outcome measures than they explain across students in academic assessment scores. This lack of explanatory power is true not only for the overall variance across both students and schools, but also for the within-school, across-student variance.

The relatively weak fit of the SEL model raises questions about how well the value-added models identify the impacts of schools on the SEL constructs and introduces the possibility that the school effect variance estimates include school-level measurement error and, thus, are overestimated. In addition, while we measure substantive differences across schools in yearly changes in SEL outcomes, the across-school variance in the average *level* of the SEL measures is proportionally much smaller than the across-school variance in the average level of the academic measures. This smaller difference in levels across schools would not be expected if substantial impacts of schools on SEL outcomes persisted and accumulated over time. Both of these findings recommend caution in interpreting measures of school differences in SEL growth as measures of the causal effects of schools on measured SEL outcomes, though they also do not rule out important school effects.

The analyses featured in this paper are a first pass at the measurement of the impacts of individual schools on SEL outcomes at a large scale, and represent the first opportunity to analyze change over time and school effects with two years of data. As additional years of SEL data from the CORE survey become available, our understanding of SEL school growth measures will improve. For example, incorporating the 2016-17 survey will make it possible to measure the stability of school SEL growth measures from one growth year (2014-15 to 2015-16) to another (2015-16 to 2016-17). In addition, research employing the results of the CORE survey will be used to improve the CORE survey itself, as part of a process of continuous improvement as in Davidson et al. (in press).

Data

The data for this study come from participating CORE districts in California. The CORE districts together serve more than one million students, nearly 20 percent of the students in California. The central dataset includes responses to surveys by students in five participating CORE districts in the spring of the 2014-15 and 2015-16 school years. The surveys include items related to four dimensions of social-emotional learning: growth mindset, self-efficacy, self-management, and social awareness. The surveys include between four and nine items related to each of the four constructs. Each item includes up to five responses indicating either the extent of a student's agreement with a statement or the extent to which a student reports an activity or experience.

The four social-emotional learning constructs are described in West, Buckley, Krachman, and Bookman (in press) 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). Finally, *social awareness* is the ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethnical norms for behavior, and to recognize family, school, and community resources (CASEL, 2005)."

Complementing the data from the student SEL survey are data from the Smarter Balanced (SBAC) assessments in math and ELA, which students in grades three through eight complete across California. The Smarter Balanced assessment is a computer-adaptive assessment aligned to the Common Core standards. The state administered these in the spring of 2014-15 and 2015-16, allowing us to compare growth in SEL to growth in math and ELA achievement. Because the SBAC is administered in the spring of grades three through eight, it is only possible to measure growth, which requires both a current outcome measure and a prior outcome measure in math and ELA, among students in grades four through eight. Given that it is possible only in these grades to compare SEL growth measures to more traditional academic growth measures, this study focuses on SEL growth in grades four through eight.

The samples used in producing the SEL growth measures are made up of students in CORE who responded to the survey in both 2014-15 and 2015-16. Students must have responded to at least half of the survey questions associated within a given SEL construct for their responses to have been considered valid. To be included in the growth measure for a given SEL construct, students must have had valid survey responses in 2015-16 for that particular construct, as well as valid responses in 2014-15 for all four constructs. Valid responses to all four constructs in 2014-15 were required because all four were included as control variables in the growth model. In addition, students must also have had SBAC scores in math and ELA in 2014-15 and demographic data available to serve as additional control variables in the growth model.

Similarly, we estimate the SBAC growth measures for a given subject using a sample of students in CORE with SBAC scores in that subject in 2015-16, with SBAC scores in both subjects and with valid responses in all four SEL constructs in 2014-15, and with available demographic data.

Table 1. Descriptive Statistics

Panel A: Demographics (self-efficacy outcome sample)

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Male	0.50	0.50	0.50	0.50	0.50
Asian	0.04	0.05	0.08	0.09	0.08
Hispanic/Latino	0.75	0.73	0.71	0.70	0.70
African American	0.07	0.07	0.07	0.06	0.07
ELL	0.37	0.29	0.25	0.17	0.15
Disability	0.10	0.11	0.11	0.10	0.10
Econ. Disadv.	0.81	0.80	0.81	0.79	0.79
Foster	0.01	0.01	0.01	0.01	0.00
Homeless	0.01	0.02	0.02	0.02	0.03

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

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Number of students					
ELA	35,633	42,520	42,565	38,917	39,334
Math	35,614	42,501	42,525	38,888	39,252
Growth mindset	31,169	37,750	33,917	31,627	31,989
Self-efficacy	31,244	37,812	33,929	31,647	32,017
Self-management	31,290	37,856	34,005	31,730	32,078
Social awareness	30,903	37,558	33,803	31,481	31,910
Number of schools					
ELA	633	711	472	283	273
Math	635	712	472	281	272
Growth mindset	546	631	355	209	202
Self-efficacy	546	631	355	209	202
Self-management	546	631	356	209	202
Social awareness	545	631	354	209	202

Table 1 describes the students in the sample. The first panel of Table 1 characterizes all students in the sample for whom growth measures in self-efficacy are available. The demographic makeup of the samples used to produce growth measures for the other three SEL constructs and for math and ELA are similar. Seventy to 75 percent of the sample is Hispanic/Latino, about seven percent is black, and four to nine percent is Asian. Approximately 80 percent is eligible for subsidized lunch. Thirty seven percent of fourth graders are English learners (ELL) but this number drops to 15 percent by eighth grade.

The second panel of Table 1 presents the number of students and schools in the sample for each grade and for each outcome variable. We have fewer students and schools for the SEL measures than for math and ELA. This smaller sample is in part a result of non-response or incomplete responses to the survey that was used to produce the SEL constructs. It is also in part a result of the districts administering two different forms of the survey; the regression sample includes SEL measures only for students to whom the more common of the two forms was administered.[2] Finally, differences in the number of students across grades is in part the result of differences in participation over the five sample districts in the SEL survey and SBAC assessment.

Measuring SEL growth using these data requires us to transform the responses to the SEL items on the student survey into a metric. We create scale scores for each of the four SEL constructs for students who responded to at least half of the survey items associated with that construct.[3] We use a generalized partial credit model (GPCM) to produce a scale score for each of the four constructs from the responses to these items (Meyer, Wang, & Rice, 2017). Based on Muraki's (1992) extension of the partial credit model (Masters, 1982), GPCM can incorporate measures for which responses are on a multipoint scale in contrast to dichotomous items.[4]

[2] A small percentage of students (about 1 percent) in the SEL growth measures are associated with CORE schools other than those in the five districts participating in the survey for the purpose of measuring growth at the school level. This is because students were linked for this purpose to schools with which they were associated in the SBAC. In addition, because the SBAC growth measures only require data regarding the SEL constructs in 2014-15 and students may move to another CORE district in 2015-16, a small percentage of students in the SBAC growth measures (less than 1 percent in all grades other than grade six; 2.6 percent in grade six) are associated with CORE schools outside of the five participating CORE districts. These schools are represented among the school growth measures, albeit with a very small number of students associated with them. Excluding the growth measures of these schools does not have a substantive effect on the growth measures' variance (as measured in Table 5) or correlations with each other (as measured in Table 6).

[3] Changing the definition of a valid response to include only students who responded to all survey items for a given construct yielded school growth measures with correlations of between 0.86 and 0.96 with the set of school growth measures used in this study. The differences between these school growth measures are in part a result of the sample becoming smaller when the more restrictive all-items criterion is applied (by a degree of between 53 percent and 72 percent). As the size of the sample becomes smaller, the extent to which the variance of measured school growth reflects randomness in growth across the remaining students becomes greater.

[4] Using a partial credit model (PCM) in place of the GPCM to produce SEL scale scores yielded very similar school growth measures, with correlations of .998 or greater depending on grade and construct. Using raw scores also yielded very similar growth measures, with correlations between .982 and .999. An important advantage of using scales based on GPCM (or the PCM), as opposed to raw scores, is that the model provides consistent estimates of scale scores for students with some missing responses, given the assumption that nonresponse to an item is random given the true individual scale – the assumption of local independence.

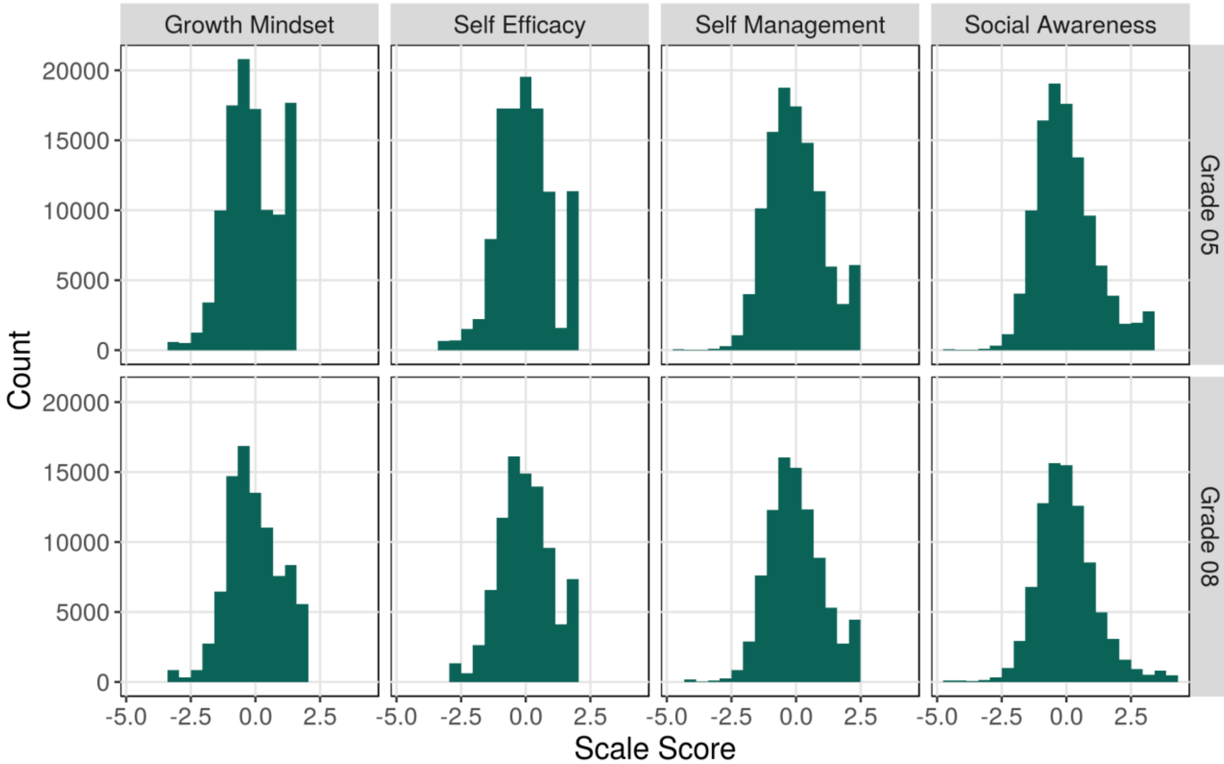
Table 2 presents the reliabilities of the SEL scale scores, measured using Cronbach's alpha. For comparison, Table 2 also presents the reliabilities of the computer-adaptive SBAC assessment in mathematics and ELA, computed using IRT conditional standard errors of measurement (SEMs). The reliabilities of the SEL measures are lower than the reliabilities of the SBAC measures, regardless of whether they are measured using Cronbach's alpha or IRT conditional SEMs. This lower reliability results, at least in part, from the small number of items used to produce the SEL measures relative to the achievement measures.

Table 2. Reliability of SEL and SBAC Scale Scores, 2014-15 and 2015-16

	Growth mindset	Self-efficacy	Self-management	Social awareness	ELA	Math
2014-15						
Grade 3	0.61	0.79	0.81	0.77	0.91	0.93
Grade 4	0.62	0.83	0.84	0.78	0.91	0.93
Grade 5	0.66	0.85	0.85	0.79	0.92	0.91
Grade 6	0.67	0.86	0.86	0.80	0.90	0.91
Grade 7	0.70	0.88	0.88	0.82	0.91	0.89
Grade 8	0.73	0.88	0.88	0.82	0.91	0.90
2015-16						
Grade 3	0.63	0.79	0.80	0.76	0.92	0.94
Grade 4	0.64	0.83	0.82	0.77	0.92	0.94
Grade 5	0.68	0.86	0.84	0.79	0.93	0.92
Grade 6	0.70	0.88	0.86	0.81	0.91	0.93
Grade 7	0.72	0.88	0.88	0.82	0.92	0.91
Grade 8	0.75	0.89	0.88	0.82	0.92	0.91
No. items	4	4	9	8	NA	NA

The distributions of the SEL scale scores exhibit some evidence of ceiling effects, with a substantial proportion of students choosing the most affirmative response to every item within a construct. Figure 1 presents histograms of the SEL scale scores in grades five and eight as examples; histograms for other grades are similar. Relatedly, the SEL raw scores display some degree of rightward skew, which is substantially mitigated in the transformation to scale scores. While the ceiling effects exhibited by the scale scores present challenges for looking at changes among individual students, they do not necessarily inhibit the school-level measures that are the focus of this paper (Koedel & Betts, 2009).

Figure 1. Histogram of Distribution of SEL Scale Scores, Grades 5 and 8



Between- and Within-School Variance and Across-Year Covariance of SEL Measures

Before proceeding to creating and assessing the measures of each school's value-added to SEL, we describe the variance in the underlying SEL measures. To the extent that schools affect SEL, we would expect variation in students' social emotional skills across schools that is relatively stable over time. To evaluate the across-school and within-school components of both the variance of the scale scores in a given year and the covariance of the scale scores from one year to the next, we estimate the following seemingly-unrelated-regressions (SUR) model:

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

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

where *j* is the school attended by student *i* in year *t*; $Y_{cij\tau}$ is measured scale score in construct *c* of student *i* in year τ; $\mu_{cj\tau}$ is the component of the variance of construct *c* in year τ that is across year-*t* schools; and $\eta_{cij\tau}$ is the component of the variance of construct *c* in year τ that is within year-*t* schools. School attended in year *t* is used to break down variance in both year *t* and year *t*-1 to parallel the construction of a school growth model, in which the average growth in scale scores between year *t*-1 and year *t* is used to estimate the impact of schools attended in year *t*.

We estimate Equations (1) and (2) using SUR, and from those results obtain 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}. Note that the variances and covariances involving the student-level $\eta_{cij\tau}$ terms include not only variance across students within schools in a construct, but also variance from randomness with which the assessment or survey measures each student's academic or SEL outcomes.

Table 3 presents the across-school and within-school variance of the academic subject and SEL construct scale scores in grades five and eight. It also presents the across-school and within-school correlation between current and lagged scale scores in the same grades. Two findings are noteworthy. First, the proportion of the variance in the SEL scale scores that is across-school is small in comparison to the same proportion of the variance in the SBAC scale scores. For example, in grade five, only four percent of the variance in the social awareness scale score is across-school, compared to 22 percent of the variance in the mathematics scale score. This smaller across school variation could be due to greater measurement error in the underlying variables or to a smaller school effect. Second, the correlation from year to year in the SEL scale scores is substantially lower than that in the SBAC scale scores for both across-school and within-school components. This pattern suggests that student-level SEL outcomes, as measured by the survey, have low persistence over time. The pattern could be a result of a substantial effect on SEL of factors experienced over the course of the year, or it could be a result of randomness in the measure itself from year to year.

Table 3. 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		
	Total	Across-school	Within-school	Total	Across-school	Within-school
Grade 5						
English language arts	1.00	0.21	0.79	0.83	0.95	0.80
Mathematics	1.00	0.22	0.78	0.85	0.93	0.83
Growth mindset	1.00	0.10	0.90	0.33	0.65	0.30
Self-efficacy	1.00	0.05	0.95	0.43	0.56	0.42
Self-management	1.00	0.07	0.93	0.53	0.84	0.51
Social awareness	1.00	0.04	0.96	0.43	0.60	0.42
Grade 8						
English language arts	1.00	0.17	0.83	0.83	0.94	0.81
Mathematics	1.00	0.19	0.81	0.84	0.95	0.82
Growth mindset	1.00	0.03	0.97	0.44	0.90	0.43
Self-efficacy	1.00	0.03	0.97	0.53	0.81	0.52
Self-management	1.00	0.05	0.95	0.55	0.89	0.53
Social awareness	1.00	0.04	0.96	0.49	0.88	0.48

Methods

Creating Growth Measures

We model the impacts of schools in academic subjects and SEL constructs using the following value-added regression model:

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

where school *j* is the school attended by student *i* in year t; y_{cijt} is outcome in construct *c* for student *i* in school *j* in year t; y_{Cijt-1} is a 1 × 6 vector of outcomes in all six subjects and constructs (math, ELA, and the four SEL constructs) for student *i* in year *t*-1; X_{ijt} is a vector of characteristics of student *i* 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. This specification has been referred to as a covariate adjustment model (McCaffrey, Lockwood, Koretz, Louis, & Hamilton, 2004) and as a dynamic ordinary least squares model (Guarino, Reckase, & Wooldridge, 2015).

We estimate equation (3) using the errors-in-variables approach described in Fuller (1987), which uses an estimate of the variance of measurement error in the right-hand-side variables to correct the sums-of-squares-and-cross-products matrix such that it reflects the variances and covariances of the variables in the model had they not been measured with error. In this application, the variance of measurement error is measured using Cronbach's alpha for lagged SEL constructs and IRT conditional standard errors of measurement for lagged SBAC scores. Given that the right-hand-side variables are the same regardless of which outcome is used as the left-hand-side variable, it makes no difference whether equation (3) is estimated separately for each equation or jointly as a system of seemingly unrelated regressions (SUR). The student characteristics (X_{ijt}) included on the right-hand-side of the regression are gender, ethnicity, English language learner, economic disadvantage, disability status, foster status, and homeless status.

We center the school fixed effect estimates from this regression to have a weighted mean of zero, with the weight equal to the number of students associated with the school in the regression sample. As a result, the school fixed effects are measured relative to the average school effect across the schools in the sample. We use these centered school effect estimates as the measures of school growth for each of the six outcomes. Both the current and lagged scale scores are rescaled to have a mean of zero and a standard deviation of one within each regression sample so that the school growth measures are measured in units of standard deviations across students of the outcome scale score measure.

Assessing Growth Measures

We use four approaches to better understand the estimated school growth measures. First, we compute goodness-of-fit measures to assess the extent to which we can model student social-emotional development. To the extent that we can predict students' SEL, we can conjecture that we are not omitting important variables that could predict student-reported SEL, correlate with school effects and, as a result, bias our estimates of school effects on SEL. We use traditional R^2 and within-school R^2 to measure model fit. Second, we describe the variance in the estimated school growth measures. If SEL growth varies across schools, then schools are more likely to be affecting students SEL development than if SEL growth does not vary across schools. We use the standard deviation of the estimated effects, and the ratio of the estimated variance of the school effect and the sample variance of the school effect estimate, to describe this variability adjusted for sampling error in the measurement of SEL. This reliability estimate adjusts for sampling error but not for other possible forms of measurement error or systematic bias. Third, we present correlations of the school growth measures across constructs. This analysis tests whether schools in which student learn more than expected in one dimension also learn more than expected in any of the other dimensions. Finally, we look at similar correlations but within constructs across grades within a school. If students appear to improve in one grade but all of this growth were simply measurement error then improvement in one grade would likely be negatively correlated with improvement in the next grade. Alternatively, if some schools are consistently better at SEL than others, then we would expect a positive correlation across grades. For each of these four analyses we look across SEL measures and also compare the SEL measures to academic achievement measures. As a specification check, we redo the analyses with growth measures that include school-aggregate measures of student characteristics as well as the student-level achievement and SEL measures.

Results

Model Coefficients and Goodness of Fit

Coefficients and goodness-of-fit measures for regression models of academic and SEL outcomes appear in Table 4. In all SEL models, the greatest coefficients are on same-construct lag. All coefficients except one are between 0.35 and 0.56. The exception is fourth-grade growth mindset, for which the coefficient is 0.23. The largest coefficients on same-construct lag are for social awareness (0.43 to 0.56), followed by self-management (0.42 to 0.50). Coefficients on same-construct lag were generally lower for growth mindset (0.23 to 0.50) and self-efficacy (0.36 to 0.46). All of these coefficients are smaller than the coefficients on the same-subject lag in models of math and ELA achievement, which range from 0.61 (seventh-grade ELA) to 0.93 (eighth-grade math).

Table 4. Coefficients and Regression Statistics

Panel A: Coefficients from models of SEL constructs, grade 4

	Current outcome (left-hand-side variable)							
	Growth mindset		Self-efficacy		Self-management		Social awareness	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Lagged outcomes								
ELA	0.20	(0.02)	0.05	(0.02)	0.12	(0.01)	0.09	(0.02)
Math	0.07	(0.01)	0.16	(0.01)	0.08	(0.01)	-0.04	(0.01)
Growth mindset	0.23	(0.01)	0.02	(0.01)	0.00	(0.01)	0.00	(0.01)
Self-efficacy	0.05	(0.01)	0.36	(0.01)	0.00	(0.01)	-0.01	(0.01)
Self-mgmt.	0.05	(0.01)	0.04	(0.01)	0.42	(0.01)	0.05	(0.01)
Social awareness	0.01	(0.01)	0.01	(0.01)	0.05	(0.01)	0.43	(0.01)
N	31,169		31,244		31,290		30,903	
R^2	0.26		0.28		0.39		0.28	
Within-school R^2	0.14		0.20		0.29		0.20	
School Fixed Effects	546		546		546		545	

Panel B: Coefficients from models of SEL constructs, grade 5

	Current outcome (left-hand-side variable)							
	Growth mindset		Self-efficacy		Self-management		Social awareness	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Lagged outcomes								
ELA	0.17	(0.01)	0.05	(0.01)	0.09	(0.01)	0.07	(0.01)
Math	0.07	(0.01)	0.12	(0.01)	0.06	(0.01)	-0.04	(0.01)
Growth mindset	0.35	(0.01)	0.11	(0.01)	0.05	(0.01)	0.07	(0.01)
Self-efficacy	0.09	(0.01)	0.39	(0.01)	-0.02	(0.01)	-0.04	(0.01)
Self-mgmt.	0.02	(0.01)	0.05	(0.01)	0.50	(0.01)	0.07	(0.01)
Social awareness	0.02	(0.01)	0.04	(0.01)	0.06	(0.01)	0.50	(0.01)
N	37,750		37,812		37,856		37,558	
R^2	0.35		0.34		0.44		0.34	
Within-school R^2	0.21		0.25		0.35		0.24	
School Fixed Effects	631		631		631		631	

Panel C: Coefficients from models of SEL constructs, grade 6

	Current outcome (left-hand-side variable)							
	Growth mindset		Self-efficacy		Self-management		Social awareness	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Lagged outcomes								
ELA	0.17	(0.01)	-0.02	(0.01)	0.07	(0.01)	0.02	(0.02)
Math	0.05	(0.01)	0.20	(0.01)	0.07	(0.01)	0.03	(0.01)
Growth mindset	0.41	(0.01)	0.06	(0.01)	0.05	(0.01)	0.04	(0.01)
Self-efficacy	0.08	(0.01)	0.41	(0.01)	-0.03	(0.01)	-0.04	(0.01)
Self-mgmt.	0.01	(0.01)	0.04	(0.01)	0.49	(0.01)	0.03	(0.01)
Social awareness	0.02	(0.01)	0.04	(0.01)	0.07	(0.01)	0.54	(0.01)
N	33,917		33,929		34,005		33,803	
R^2	0.38		0.34		0.43		0.35	
Within-school R^2	0.27		0.28		0.34		0.26	
School Fixed Effects	355		355		356		354	

Panel D: Coefficients from models of SEL constructs, grade 7

	Current outcome (left-hand-side variable)							
	Growth mindset		Self-efficacy		Self-management		Social awareness	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Lagged outcomes								
ELA	0.11	(0.02)	-0.02	(0.02)	0.04	(0.01)	-0.01	(0.02)
Math	0.10	(0.02)	0.16	(0.02)	0.09	(0.01)	0.05	(0.02)
Growth mindset	0.48	(0.01)	0.09	(0.01)	0.04	(0.01)	0.05	(0.01)
Self-efficacy	0.07	(0.01)	0.45	(0.01)	-0.01	(0.01)	-0.05	(0.01)
Self-mgmt.	0.03	(0.01)	0.05	(0.01)	0.50	(0.01)	0.06	(0.01)
Social awareness	-0.01	(0.01)	0.04	(0.01)	0.09	(0.01)	0.56	(0.01)
N	31,627		31,647		31,730		31,481	
R^2	0.41		0.38		0.43		0.38	
Within-school R^2	0.29		0.32		0.34		0.28	
School Fixed Effects	209		209		209		209	

Panel E: Coefficients from models of SEL constructs, grade 8

	Current outcome (left-hand-side variable)							
	Growth mindset		Self-efficacy		Self-management		Social awareness	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Lagged outcomes								
ELA	0.14	(0.02)	-0.01	(0.02)	0.08	(0.02)	0.05	(0.02)
Math	0.03	(0.02)	0.19	(0.02)	0.06	(0.02)	0.01	(0.02)
Growth mindset	0.50	(0.01)	0.11	(0.01)	0.06	(0.01)	0.07	(0.01)
Self-efficacy	0.06	(0.01)	0.46	(0.01)	0.01	(0.01)	-0.03	(0.01)
Self-mgmt.	0.03	(0.01)	0.04	(0.01)	0.48	(0.01)	0.05	(0.01)
Social awareness	0.04	(0.01)	0.03	(0.01)	0.09	(0.01)	0.55	(0.01)
N	31,989		32,017		32,078		31,910	
R^2	0.41		0.41		0.41		0.38	
Within-school R^2	0.30		0.35		0.34		0.29	
School Fixed Effects	202		202		202		202	

Panel F: Coefficients from models of ELA and math, grades 4 and 5

	Current outcome (left-hand-side variable)							
	ELA, grade 4		Math, grade 4		ELA, grade 5		Math, grade 5	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Lagged outcomes								
ELA	0.68	(0.01)	0.07	(0.01)	0.70	(0.01)	0.11	(0.01)
Math	0.16	(0.01)	0.80	(0.01)	0.16	(0.01)	0.78	(0.01)
Growth mindset	0.01	(0.01)	0.01	(0.01)	0.03	(0.01)	0.01	(0.01)
Self-efficacy	0.01	(0.01)	0.02	(0.01)	0.01	(0.01)	0.02	(0.00)
Self-mgmt.	0.06	(0.01)	0.04	(0.01)	0.05	(0.01)	0.03	(0.01)
Social awareness	-0.04	(0.01)	-0.04	(0.01)	-0.03	(0.01)	-0.03	(0.01)
N	35,633		35,614		42,520		42,501	
R^2	0.82		0.85		0.84		0.86	
Within-school R^2	0.71		0.75		0.74		0.76	
School Fixed Effects	633		635		711		712	

Panel G: Coefficients from models of ELA and math, grades 6 and 7

	Current outcome (left-hand-side variable)							
	ELA, grade 6		Math, grade 6		ELA, grade 7		Math, grade 7	
	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.	Coeff.	S.E.
Lagged outcomes								
ELA	0.71	(0.01)	0.16	(0.01)	0.61	(0.01)	0.03	(0.01)
Math	0.12	(0.01)	0.70	(0.01)	0.26	(0.01)	0.89	(0.01)
Growth mindset	0.02	(0.01)	0.00	(0.01)	0.04	(0.01)	0.01	(0.00)
Self-efficacy	0.02	(0.00)	0.03	(0.00)	0.01	(0.00)	0.02	(0.00)
Self-mgmt.	0.06	(0.01)	0.05	(0.01)	0.04	(0.01)	0.03	(0.01)
Social awareness	-0.04	(0.01)	-0.05	(0.01)	-0.03	(0.01)	-0.02	(0.01)
N	42,565		42,525		38,917		38,888	
R^2	0.82		0.85		0.83		0.92	
Within-school R^2	0.73		0.76		0.73		0.81	
School Fixed Effects	472		472		283		281	

Panel H: Coefficients from models of ELA and math, grade 8

	Current outcome (left-hand-side variable)			
	ELA, grade 8		Math, grade 8	
	Coeff.	S.E.	Coeff.	S.E.
Lagged outcomes				
ELA	0.66	(0.01)	0.01	(0.01)
Math	0.24	(0.01)	0.93	(0.01)
Growth mindset	0.04	(0.00)	0.00	(0.00)
Self-efficacy	-0.01	(0.00)	0.01	(0.00)
Self-mgmt.	0.02	(0.00)	0.03	(0.01)
Social awareness	0.01	(0.00)	0.00	(0.01)
N	39,334		39,252	
R^2	0.85		0.92	
Within-school R^2	0.76		0.78	
School Fixed Effects	273		272	

Note that all regressions include lagged outcomes in ELA, math, growth mindset, self-efficacy, self-management, and social awareness; indicators for gender, economic disadvantage, English language learner (beginning, intermediate, advanced, and level not measured), foster child, homeless, race/ethnicity (Asian, African American, Hispanic/Latino), and disability (moderate, severe); and school fixed effects.

The goodness-of-fit measures are smaller in the models of the SEL measures than in the models of academic achievement. The overall R^2 of the SEL models ranges from 0.26 (fourth-grade growth mindset) to 0.44 (fifth-grade self-management), considerably lower than that of the academic subject models, which ranges from 0.82 (sixth-grade ELA) to 0.92 (eighth-grade math). Among the SEL models, the overall R^2 is higher in self-management (between 0.39 and

0.44) than in growth mindset (0.26 to 0.41), self-efficacy (0.28 to 0.41), or social awareness (0.28 to 0.38).

The R^2 measures described above include not only the explanatory power of student characteristics, such as lagged achievement, lagged SEL-related skills, and demographics, but also the explanatory power of the school fixed effects. We can also compute a measure of fit that does not include the explanatory power of the school effects, which helps to assess the extent to which the controls for prior outcomes and student characteristics predict outcomes sufficiently well to identify the effects of schools. This *within-school* R^2 measure uses within-school variation in the place of overall variation in the outcome, prediction, and error terms. As is the case in overall goodness-of-fit, the within-school measure is substantially lower in the SEL models than in the models of academic achievement. Across the four SEL outcomes and five grades, within-school R^2 ranges between 0.14 (fourth-grade growth mindset) and 0.35 (eighth-grade self-efficacy). In contrast, the same combination of explanatory variables explains about three-quarters of the variation in math and ELA achievement.

The weak fit in the SEL models raises questions about the consistency of the measured SEL school effects. A possible reason that the fit of the value-added model may be weak is the presence of substantial unobserved non-school factors that contribute to student SEL outcomes and that are not controlled for by the covariates in the model. These non-school factors, if correlated with school assignment, would introduce bias into the estimated school effects. Another possible reason for a weak fit is that the student SEL measures based on survey responses could be measuring not only SEL outcomes but also other variables, such as a student's mood at the time of taking the survey, which may not be persistent from one year to the next. This is a kind of measurement error that would not be identified as such by Cronbach's alpha, which measures internal consistency. If this measurement error is correlated with school assignment, this would also introduce bias into the estimated school effects. A third possibility is that SEL outcomes among third- to eighth-grade students are, on average, not especially persistent from one year to the next. Under this scenario, it would be possible for the school effect estimates produced by the value-added model to have little bias even if the explanatory power of the model is low. A low overall persistence of SEL outcomes would not necessarily imply that all aspects of SEL outcomes diminish substantially over time; for example, it is possible that SEL outcomes have both a transitory and a persistent component, which would make it possible for some childhood SEL interventions to have impacts on adult outcomes even if SEL persistence is low on average.

Variance and Reliability of School Growth Estimates

We estimate the overall magnitude of the impacts of schools with the noise-corrected variance of school effects in the SEL growth models as follows:

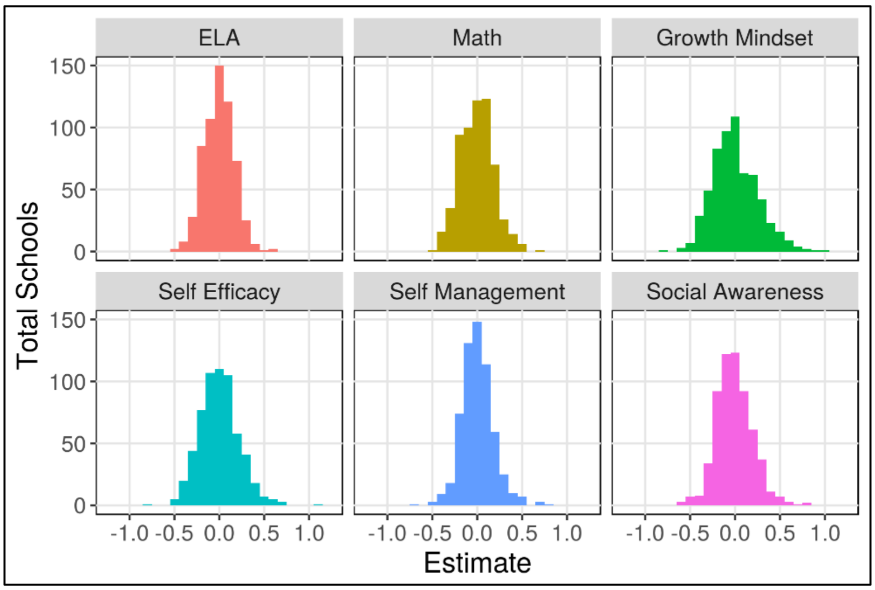
$$Est.Variance\left[\alpha_{cjt}\right] = Sample\ Variance\left[\hat{\alpha}_{cjt}\right] - Sample\ Mean\left[\hat{\sigma}_{cjt}^2\right] \quad (4)$$

where α_{cjt}, as in the notation of equation (3), are the effects for school *j* in construct *c* at time *t*; $\hat{\alpha}_{cjt}$ are the estimated school effects produced by the value-added regression and centered to have a mean of zero; and $\hat{\sigma}^2_{cjt}$ are the squares of the standard error estimates of those estimated and centered effects. This approach estimates the variance of the component of the school effects α_{cjt} in a way that does not include variance due to the estimation error of $\hat{\alpha}_{cjt}$.

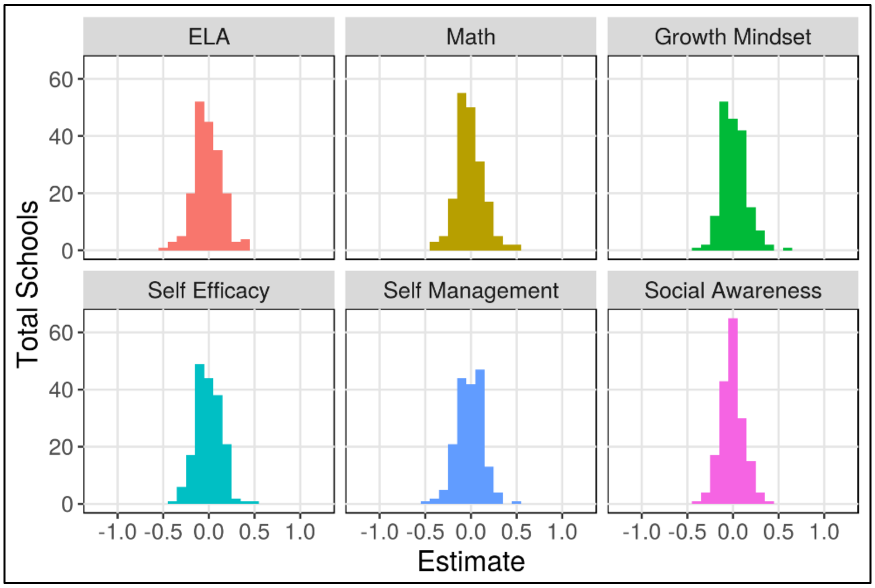
Figure 2 presents histograms of the estimated school effects for each of the six outcomes for grades five and eight; results in other grades are similar. The histograms illustrate that the range of growth measures across schools is very similar for the four SEL constructs in comparison to the two academic subjects. Moreover, the distributions are approximately normal.

Figure 2. Histograms of School Effects

Panel A: Grade 5



Panel B: Grade 8



Estimates of the noise-corrected standard deviation of the school effect in models of each of the six outcomes are presented in Table 5. Table 5 also gives the average reliability of the school effect estimates, which is the ratio of the estimated (noise-corrected) variance of the school effects α_{cjt} (the left-hand term of equation 4) and the sample variance of the school effect estimates $\hat{\alpha}_{cjt}$ (the first term on the right-hand-side of equation 4).

Table 5. Standard Deviation (Noise-Corrected) and Reliability of School Growth Effect Estimates

	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Standard deviation					
ELA	0.15	0.14	0.18	0.13	0.13
Math	0.17	0.16	0.16	0.11	0.13
Growth mindset	0.24	0.23	0.15	0.10	0.09
Self-efficacy	0.18	0.19	0.12	0.10	0.11
Self-mgmt.	0.16	0.13	0.13	0.11	0.11
Social awareness	0.16	0.16	0.13	0.11	0.10
Reliability					
ELA	0.85	0.84	0.93	0.93	0.91
Math	0.87	0.87	0.91	0.91	0.91
Growth mindset	0.82	0.80	0.75	0.75	0.62
Self-efficacy	0.71	0.75	0.69	0.69	0.74
Self-mgmt.	0.68	0.62	0.72	0.72	0.76
Social awareness	0.67	0.65	0.67	0.67	0.67

The results in Table 5 show that the variance of school effects in models of growth in SEL constructs is similar to the variance of school effects in models of growth in academic subjects.

For example, in fifth grade, the estimated standard deviation across schools of effects on growth mindset is equal to 0.23 times the standard deviation of growth mindset outcomes across students. Across all five grades, the estimated standard deviation of school effects is in the range of 0.09 and 0.24 across the four SEL constructs and in the range of 0.11 and 0.18 in math and ELA. The standard deviation tends to be lower in middle school grades than in elementary school grades across all four SEL constructs and both academic subjects.

The similarity in school variance between SEL and academic measures could suggest that schools impact SEL outcomes. However, the estimated variance of school effects may also be the result of bias. This potential bias is relevant if, as discussed earlier, there are substantial non-school effects on SEL outcomes that are not controlled for by the model's covariates and that are correlated with school assignment, or if the SEL measures are measuring not only SEL outcomes but also other non-SEL variables that are correlated with schools. This bias, unless it is substantially negatively correlated with the actual impacts of schools, would lead to overestimation of the variance.

Correlations of School Growth Measures Across Constructs

Table 6 presents the correlations between school growth measures across the two academic and four SEL outcomes. The correlations within the academic measures and within the SEL measures are larger than those between the academic and SEL measures. These differences are not unexpected. Prior work on teachers has shown that teachers who excel in improving students' academic achievement do not necessarily improve other outcomes for students (Jackson, in press; Liu and Loeb, 2018), and the same logic may apply to school effects.

Table 6. Correlation of Academic and SEL School Growth Measures

	ELA	Math	Growth mindset	Self-efficacy	Self-management	Social awareness
Grade 4						
ELA	1.00	0.68	0.17	0.08	0.15	0.15
Math	0.68	1.00	0.16	0.13	0.19	0.14
Growth mindset	0.17	0.16	1.00	0.40	0.26	0.38
Self-efficacy	0.08	0.13	0.40	1.00	0.55	0.61
Self-management	0.15	0.19	0.26	0.55	1.00	0.67
Social awareness	0.15	0.14	0.38	0.61	0.67	1.00
Grade 5						
ELA	1.00	0.63	0.23	0.13	0.16	0.18
Math	0.63	1.00	0.20	0.09	0.14	0.13
Growth mindset	0.23	0.20	1.00	0.49	0.30	0.38
Self-efficacy	0.13	0.09	0.49	1.00	0.55	0.54
Self-management	0.16	0.14	0.30	0.55	1.00	0.61
Social awareness	0.18	0.13	0.38	0.54	0.61	1.00
Grade 6						
ELA	1.00	0.74	0.38	0.15	0.26	0.25
Math	0.74	1.00	0.42	0.15	0.26	0.26
Growth mindset	0.38	0.42	1.00	0.25	0.20	0.31
Self-efficacy	0.15	0.15	0.25	1.00	0.51	0.48
Self-management	0.26	0.26	0.20	0.51	1.00	0.70
Social awareness	0.25	0.26	0.31	0.48	0.70	1.00
Grade 7						
ELA	1.00	0.43	0.29	0.04	0.22	0.25
Math	0.43	1.00	0.18	0.08	0.12	0.23
Growth mindset	0.29	0.18	1.00	0.37	0.34	0.35
Self-efficacy	0.04	0.08	0.37	1.00	0.47	0.43
Self-management	0.22	0.12	0.34	0.47	1.00	0.68
Social awareness	0.25	0.23	0.35	0.43	0.68	1.00
Grade 8						
ELA	1.00	0.42	0.21	0.18	0.23	0.27
Math	0.42	1.00	0.23	0.12	0.18	0.30
Growth mindset	0.21	0.23	1.00	0.29	0.27	0.18
Self-efficacy	0.18	0.12	0.29	1.00	0.47	0.34
Self-management	0.23	0.18	0.27	0.47	1.00	0.64
Social awareness	0.27	0.30	0.18	0.34	0.64	1.00

Correlations of School Growth Measures Within Schools Across Grades

Table 7 presents the correlations of school growth measures across grades within schools. Only correlations within elementary grades and within middle grades are presented, with sixth grade included as both an elementary and a middle grade. Fewer than fifty schools in the CORE sample included both elementary and middle grades.

Table 7. Correlations of School Effects Across Grades Within Schools

	Elementary grades			Middle grades		
	4 and 5	5 and 6	4 and 6	6 and 7	7 and 8	6 and 8
ELA	0.20	0.15	0.23	0.24	0.40	0.18
Math	0.10	0.12	0.06	0.00	0.28	0.20
Growth mindset	0.17	0.05	0.05	0.26	0.30	0.12
Self-efficacy	0.18	0.05	0.12	0.22	0.28	0.24
Self-mgmt.	0.09	0.04	0.11	0.33	0.47	0.32
Soc. awareness	0.09	-0.02	-0.16	0.27	0.37	0.37

The correlations among school SEL growth measures across grades are modest, but they are similar to the correlations across grades among the academic growth measures; the average of the correlations presented in Table 7 is 0.18 among the SEL constructs, and also 0.18 among the academic subjects. These modest correlations suggest that there is substantial variation in the impacts of schools on both academic and SEL outcomes by grade. This pattern would be consistent with the presence of variance in effects across individual teachers and classrooms within schools. The within-school, across-grade correlation is generally greater in the middle grades than in the elementary grades for both the SEL and academic outcomes; this result potentially stems from teachers teaching in multiple grades within middle schools and students experiencing multiple teachers each year.

Sensitivity of School Growth Measures to Including School-Level Covariates

The value-added model employed throughout this paper includes only student-level variables among its covariates. An alternative value-added model controls not only for student-level variables, but also for school-level variables. One version of this model, which controls for school-level averages of all variables included as student-level covariates, is described in equations (3a) and (3b):

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

$$\alpha_{cjt} = \zeta_c + \theta_c \bar{y}_{Cjt-1} + \bar{X}_{jt}\varphi_c + u_{cjt} \qquad (3b)$$

where (3a) is the same as (3), $\bar{y}_{Cjt-1}$ is a vector of average prior academic and SEL outcomes at school *j*, $\bar{X}_{jt}$ is a vector of average student demographics at school *j*, and u_{cjt} is the value-added effect of school *j*. The model expressed in (3a) and (3b) has the advantage of controlling for any school-level peer effects that are correlated with average prior academic and SEL outcomes or demographics. It has the disadvantage of partialling out from value-added any aspect of school quality that is correlated with these school-level averages. The model can be estimated using a two-step approach. First, the regression in (3a) is estimated using an errors-in-variables regression, as described earlier in the discussion of equation (3), to produce the school fixed effects estimates $\hat{\alpha}_{cjt}$. Second, the school fixed effects estimates $\hat{\alpha}_{cjt}$ are regressed on the school-level averages $\bar{y}_{Cjt-1}$ and $\bar{X}_{jt}$ by ordinary least squares as in (3b). The residual from the second regression are estimates of the value-added effects u_{cjt}. Alternatively, the model can be estimated in a single step by regressing the school-level averages of the left-hand-side variable in (3a) on $\bar{y}_{Cjt-1}$ and $\bar{X}_{jt}$ using ordinary least squares. The residuals from this regression will estimate the value-added effects u_{cjt}.

Table 8 presents correlations between estimated value-added measures between the model used in the bulk of this paper, which only controls for student-level covariates, and a model that also controls for school-level averages of the student-level covariates. For the most part, the correlations are high. Even when correlations are high, however, the value-added measures of individual schools may differ substantially between the two approaches.

Table 8. Correlations Between Value-added Measures that Include and Do Not Include Controls for School-level Averages of Student-level Covariates

Grade	4	5	6	7	8
ELA	0.95	0.99	0.96	0.89	0.93
Mathematics	0.95	0.97	0.97	0.92	0.95
Growth mindset	0.98	0.98	0.95	0.89	0.90
Self-efficacy	0.95	0.93	0.94	0.93	0.86
Self-management	0.97	0.96	0.91	0.85	0.85
Social awareness	0.98	0.97	0.94	0.86	0.85

The properties of the SEL school growth measures produced by the model that controls for school averages are, for the most part, similar to that of the model that controls for student-level variables only. The fit of the model is not substantially improved by the inclusion of the school-level averages.[5] The estimated variances of both the academic and SEL school effects are lower in the model that controls for school-level averages, which is an expected result given that the component of school effects that is correlated with school averages is partialled out. However, the variances are not lower in a way that is sufficiently disproportionate to change the result that the variance of the school effects in the SEL constructs is of a similar magnitude

[5] Both overall R-squared and within-school R-squared are algebraically the same between the two models, and so neither measure sheds much light on the relative fit of the two models. An alternative measure of fit compares the variance of the explained component minus the school effect $\hat{\xi}_c + \hat{\lambda}_c y_{Cijt-1} + \hat{\beta}_c X_{ijt} + \hat{\theta}_c \bar{y}_{Cjt-1} + \hat{\varphi}_c \bar{X}_{jt}$ to the variance of the error component plus the school effect $\hat{u}_{cjt} + \hat{\varepsilon}_{cijt}$. The ratio of these two is not substantially different from the analogous ratio in the model that does not control for school effects among the SEL construct or academic subject models.

to the variance of the school effects in academic subjects. After controlling for school averages, the estimated standard deviations of school effects are in the range of 0.06 to 0.23 among the SEL constructs and in the range of 0.10 to 0.17 among the academic subjects.

One substantive difference between the properties of the growth measures produced by the two models is the correlations among effects within schools across the middle grades. We present a comparison of these correlations in Table 9. The correlations are somewhat lower in the model that includes school averages, especially for self-management and social awareness, suggesting that a part of the correlation among estimated effects within schools across the middle grades is driven by the component of those effects that is correlated with observable student characteristics.

Table 9. Correlations Between Middle Grades within Schools between Models that Control and Do Not Control for School Averages

	Model without school averages			Model with school averages		
	6 and 7	7 and 8	6 and 8	6 and 7	7 and 8	6 and 8
ELA	0.24	0.40	0.18	0.27	0.43	0.22
Math	0.00	0.28	0.20	0.03	0.24	0.15
Growth mindset	0.26	0.30	0.12	0.23	0.20	0.09
Self-efficacy	0.22	0.28	0.24	0.25	0.18	0.14
Self-mgmt.	0.33	0.47	0.32	0.21	0.33	0.22
Soc. awareness	0.27	0.37	0.37	0.14	0.22	0.26

Conclusion

Using data from a large-scale survey panel of more than 150,000 students in five California districts that includes items measuring SEL outcomes, we produced and evaluated measures of the impacts of individual schools on social-emotional outcomes by grade. To our knowledge, this is the first attempt to produce school growth measures of SEL outcomes at a large scale.

The student surveys include items relevant to four SEL outcomes: growth mindset, self-efficacy, self-management, and social awareness. We use measures of these four SEL outcomes based on responses to the student survey as outcome variables in value-added growth models for grades four through eight. We estimate the value-added models using linear regressions of current SEL outcome on lagged SEL outcomes, lagged math and ELA achievement, student demographics, and school fixed effects, with control for measurement error in all lagged SEL and achievement measures. The specification of this value-added growth model is similar to that often used to measure the impacts of schools in academic subjects such as math and ELA, which we also estimate for schools in the districts administering the survey.

We find variance across schools in measured impacts on SEL outcomes that is similar to the estimated variance across schools in impacts on academic outcomes. Across the four SEL outcomes and five grades covered by this study, we estimated a standard deviation of school effects in the range of 0.09 and 0.24 times the standard deviation of the level of SEL outcome measures across students. The analogous standard deviation estimates in models of math and ELA were in the range of 0.11 to 0.18.

However, the fit of the value-added models of SEL outcomes is relatively weak compared to value-added models of mathematics and English language arts. While the covariates in the value-added models explain about three-quarters of the variation in math and ELA achievement across students within schools, they do not typically explain more than a third of the variation in the SEL measures. The lack of explanatory power suggests that either measurement error in the outcome or omitted variables bias may be a concern. In addition, the SEL measures have more variation within schools than between schools relative to the academic measures, which again suggestions a concern for measurement error. While we can adjust for sampling error in measurement we cannot adjust for other factors that might affect students survey responses. As a result, while the variance of the school effect estimates is promising, we recommend interpreting these results with caution as measures of the causal effects of schools on students' social emotional learning.

The school SEL growth measures described in this study are based on two years of CORE student survey data about student SEL outcomes—the minimum sufficient for measuring growth, and, to our knowledge, the first panel data set of this size of its kind. As more years of data become available, it will become possible to explore additional issues, including the stability of SEL growth measures for individual schools from year to year, as well as to continue to explore the potential to distinguish the effects of schools on students' SEL in further depth. In addition, continued research on SEL outcomes will inform the evolving design of the CORE survey and of the SEL measures, which will affect the school SEL growth measures in turn. Given the newness of the data, it is most appropriate to understand these results as a first pass at understanding the potential for measuring the impacts of individual schools on SEL outcomes.

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.

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.

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78*(1), 246–263

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis*, *39*(1), 146-170.

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.

Briggs, D., & Weeks, J. (2011). The persistence of school-level value-added. *Journal of Educational and Behavioral Statistics, 36*(5), 616-637.

Chamberlain, G. E. (2013). Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of Sciences of the United States of America, 110*(43), 17176–82.

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.

Cohen, G.L., Garcia, J., Apfel, N., & Master, A. (2006.) Reducing the racial achievement gap: A social-psychological intervention. *Science, 313*, 1307–1310.

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: Author.

Davidson, L., Crowder, M., Gordon, R., Domitrovich, C., Brown, R., & Hayes, B. (in press). A continuous improvement approach to social and emotional competency measurement. *Journal of Applied Developmental Psychology*.

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

Durlak, J., Weissberg, R., Dymnicki, A., Taylor, R., & Schellinger, K. (2001). 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.

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.

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

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.

Hallinger, P. (2005). Instructional leadership and the school principal: A passing fancy that refuses to fade away. *Leadership and Policy in Schools, 4*(3): 221–39.

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

Jackson, K. (In press). What do test scores miss? The importance of teacher effects on non-test-score outcomes. *Journal of Political Economy*.

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. (2002). 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.

Kautz, T., Heckman, J., Diris, R., ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. Paris: Organization for Economic Co-Operation and Development.

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.

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.

Ladd, H., & Walsh, R. (2002). Implementing value-added measures of school effectiveness: Getting the incentives right. *Economics of Education Review, 21*(1), 1-17.

Leithwood, K., Louis, K. S., Anderson, S., & Wahlstrom, K. (2004). *Review of research: How leadership influences student learning.* St. Paul, MN: Center for Applied Research and Educational Improvement.

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

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

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., Wang, Y., & Rice, A. (2017). *Measuring students' social-emotional learning among California's CORE districts: An IRT modeling approach*. Paper presented at the 39th Fall Research Conference of the Association for Public Policy Analysis and Management, Chicago, IL.

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

Olweus, D. (1994). Bullying at school. In: L.R. Huesmann (Ed.) *Aggressive Behavior: Current Perspectives (*pp. 97-130). Boston, MA: Springer.

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.

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.

Thomsen, J. (2013). *50-State comparison: State school accountability report cards*. Retrieved from https://www.ecs.org/state-school-accountability-report-cards/

Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology 7*(27), 27-56.

Waters, T., Marzano, R. J., & McNulty, B. (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Aurora, CO: Mid-Continent Research for Education and Learning.

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