Measuring Social-emotional Learning Over Time:
Implications for Education and Evaluation

Social-emotional learning (SEL) is an old concept that is gaining new traction in education practice and policy. One reason for the renewed interest in SEL is research providing evidence on the importance of social-emotional competencies to long-term outcomes like high school graduation and earnings (Almlund, Duckworth, Heckman, & Kautz, 2011; Belfield et al., 2015; Dweck, Walton, & Cohen, 2011; Heckman & Vytlacil, 2001). This interest has manifested itself in policy and practice. For example, a consortium of California districts serving over one million students included SEL measures in its No Child Left Behind waiver. Additionally, The Every Student Succeeds Act (ESSA) of 2015 requires states to include non-academic indicators in their accountability plans.

Despite the growing emphasis on SEL, there are still fundamental questions about measuring these competencies and using them for evaluation purposes that remain unanswered. In particular, little is known about how SEL competencies develop for students over time, and how teachers and schools contribute to that development. These gaps in the literature have arisen in part because few available datasets track individual student scores on SEL constructs over time. Data that do include longitudinal SEL measures often come from federal datasets like the Early Childhood Longitudinal Study (ECLS), which have limitations in terms of grades, constructs covered, and recency.

Thanks to emerging sources of data, the four papers in this symposium begin to close that gap in the research by modeling SEL scores over time and estimating teacher and school contributions to them. The first two papers estimate teacher effects on SEL, including whether those effects persist into later grades. One of these papers shows that teacher effects on students' happiness in class has positive effects on later behavior, resulting in decreased numbers of out-of-school suspensions three years later. The other estimates teacher contributions to students' growth for four SEL-related constructs.

Meanwhile, the third and fourth paper use structural equation models to better understand how constructs like academic self-efficacy and growth mindset develop over time, and how much of that development is attributable to schools. The first study examines the state-trait balance of associated constructs over a three-year period, showing that less than half the variance in scores is stable over time. This finding suggests that SEL constructs are likely driven more by contextual and environmental factors than enduring student characteristics, which has implications for how teachers and schools should attempt to foster SEL. The last study jointly models math self-efficacy and achievement to show how the two influence each other's development. Initial results suggest that models fit better when high self-efficacy at one time point is associated with higher achievement test scores at a later time point rather than vice-versa, though this process does not appear dependent on the school context.

Below we provide detail on each paper, including the ways in which they contribute to our understanding of how SEL competencies develop over time.

Paper 1
Persistence of Teacher Effects on Students' Academic and Non-Academic Skill Formation

David Blazar, University of Maryland

1. Background and Context

Decades worth of research on education production have narrowed in on the importance of teachers to students' skill formation (Murnane and Phillips 1981; Todd and Wolpin 2003). Recognizing the multidimensional nature of student knowledge, more recent work also has begun to show that teachers can influence outcomes beyond test scores, including self-reported measures of behavior, self-efficacy, and engagement (Blazar, 2018; Jennings and DiPrete 2010; Kraft forthcoming) and observed school behaviors such as absences and suspensions (Gershenson 2016; Jackson 2012). However, teacher effects on academic versus non-academic outcomes generally are unrelated to each other – sometimes even negatively related (Blazar 2018). These findings highlight a pressing need to develop teacher skill in a way that is consistent with and supports students' academic *and* non-academic skill formation over the life cycle.

2. Purpose and Objectives

I extend this growing body of work on the multidimensional effect of teachers on students in two keys ways. First, I estimate the longer-term effects of teachers on students' academic and non-academic skills. Specifically, I track the effect of teachers that students had in fourth or fifth grade through middle school and into the beginning of high school, thus estimating the degree of persistence versus decay in teacher effect estimates. Second, I extend prior work by examining the extent to which observed classroom behaviors predict longer-term student outcomes. Observations of teaching practice are more ideal for tracking and improving teacher behavior than "value-added" scores, which simply rank teachers with regard to their ability to improve student skill and so cannot tell us what teachers do or the supports they need to produce stronger student outcomes.

3. Data and Sample

This paper draws on a combination of district administrative records and data from the National Center for Teacher Effectiveness (NCTE) project. The NCTE project was conducted over the course of three school years (2010-11 through 2012-13) in three districts on the East coast of the United States. The project collected measures of students' non-academic outcomes through a survey administered in the spring of each school year: *Behavior in Class* (internal consistency reliability [α] is 0.74), *Self-Efficacy in Math* (α = 0.76), and *Happiness in Class* (α = 0.82). The study also collected videotapes of teachers' mathematics lessons, providing direct measures of teacher behavior and classroom practice. These math lessons were scored by trained raters on two observation instruments: the Classroom Assessment Scoring System (CLASS) and the Mathematical Quality of Instruction (MQI) that together capture four unique teaching constructs (see Blazar et al. 2017 for factor analyses): *Emotional Support*, *Classroom Organization*, *Ambitious Mathematics Instruction*, and *Mathematical Errors*.

While the NCTE-collected data allow me to capture rich measures of teacher effectiveness in the short term, the district administrative data allow me to examine the predictive power of these measures on students' skill formation over time. District administrative records – which were available during the NCTE study beginning in 2010-11 and all subsequent years through 2016-17 – include demographic data, test scores in math and reading, observed school behaviors (i.e., absences, suspensions), and, in one district, student-reported measures of their engagement.

4. Experimental Design

In the spring of 2012, the NCTE project team worked with staff at participating schools to randomly assign sets of teachers to class rosters of the same grade level that were constructed by principals or school leaders. To be eligible for randomization, teachers had to work in schools and grades in which there was at least one other participating teacher, and their principal had to consider these teachers as capable of teaching any of the rosters of students designated for the group of teachers.

5. Research Design and Analysis

All analyses begin with a standard model of skill production (Todd and Wolpin 2003) for student *i* in district *d*, school *s*, and grade *g*, taught by teacher *j* in year *t*:

$$OUTCOME_{idsgj} = \alpha f(A_{it-1}) + \pi X_{it} + \varphi \bar{X}_{it}^{c} + \nu_{sg} + (\tau_{jt} + \varepsilon_{idsgjt}) \quad (1)$$

$OUTCOME_{idsgjct}$ is used interchangeably for each academic or non-academic outcome, which is specified as a function of prior academic performance, A_{it-1}; observable student, X_{it}, and class, $\bar{X}_{it}^{c}$, characteristics; and the school-grade, ν_{sg}, they were in in a given year. Because of the randomized design, I can estimate the causal effect of teachers on student outcomes, τ_{jt}.

Equation (2) estimates the relationship between these teacher effect estimates and longer-term student outcomes:

$$OUTCOME_{idsgj,t+n} = \delta \hat{\tau}_{jt=2012-13} + \kappa_{sgt=2012-13} + \varepsilon_{idsgjt} \quad (2)$$

I capture outcomes in all years *t* + *n*, where *t* refers to the year in which students participated in the NCTE study and *n* refers to the number of subsequent years I am able to track students. My main parameters of interest are in the vector, δ, which describe the relationship between current student outcomes captured in middle school and the beginning of high school and several characteristics of students' fourth- or fifth-grade teachers.

5. Results

Experimental teacher effects captured in fourth or fifth grade also predict longer-term student outcomes up to four years later. Having a fourth or fifth grade teacher who is effective at improving math test scores increase math achievement in later years, though this relationship falls from 1:1 in the year after having that teacher to roughly 0.5 SD several years later. Teacher effects on students' *Happiness in Class* also have positive effects on later behavior, resulting in decreased number of days of out-of-school suspensions three years later.

In order to unpack what it is that effective teachers do to improve student skill, I examine relationships between observed teaching practices and longer-term student outcomes. I find that higher scores on *Ambitious Mathematics Instruction* results in fewer days absent two years later. Having a teacher with high scores on *Classroom Organization* improves math test scores up to two years later, and decreases the number of days absent from school four through six years later.

6. Conclusion

For many, these results may seem intuitive: teachers impact students in myriad ways that have lasting impacts on their skill development. These findings, though, can motivate increased policy attention to the complexities of – as well as likely complementarities between – student and teacher skill formation.

References

Blazar, David. 2018. Validating teacher effects on students' attitudes and behaviors: Evidence from random assignment of teachers to students. *Education Finance and Policy* 13(3): 281-309.

Blazar, David, David Braslow, Charalambos Y. Charalambous, and Heather C. Hill. 2017. Attending to general and mathematics-specific dimensions of teaching: Exploring factors across two observation instruments. *Educational Assessment* 22(2): 71-94.

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

Jackson, C. Kirabo. 2012. Non-cognitive ability, test scores, and teacher quality: Evidence from ninth grade teachers in North Carolina. NBER Working Paper 18624.

Jennings, Jennifer L., and Thomas A. DiPrete. 2010. Teacher effects on social and behavioral skills in early elementary school. *Sociology of Education* 83(2): 135-159.

Kraft, Matthew A. Forthcoming. Teaching effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*.

Murnane, Richard J., and Barbara R. Phillips. 1981. What do effective teachers of inner-city children have in common? *Social Science Research* 10(1): 83-100.

Todd, Petra E., and Kenneth I. Wolpin. 2003. On the specification and estimation of the production function for cognitive achievement. *The Economic Journal* 113(485): F3-F33.

Paper 2
Measuring Teachers' Impacts on Students' Social Emotional Development

Robert Meyer, Education Analytics
Andrew Rice, Education Analytics
Libby Pier, Education Analytics

1. Background and Context

Increasingly, educators, families, policymakers, and researchers are recognizing that students' social-emotional learning (SEL) is central to ensuring their success, not only in terms of academic outcomes, but also in terms of longer-term life outcomes (Nagaoka et al., 2015). Research demonstrates that educators can support the development of these "non-cognitive" skills, both directly via SEL programs led by teaching staff (Durlak, Dymnicki, Taylor, Weissberg, & Schellinger, 2011) and indirectly through broader policies and practices that improve a school's culture and climate (McCormick, Cappella, O'Connor, & McClowry, 2015). However, the ability to assess the impact of such programs or policies relies on reliably measuring students' social-emotional skills. There is a growing body of innovative and promising tools aimed at assessing students' SEL (McKown & Taylor, 2018), but these assessments have not been administered to large samples of students across multiple grades and from diverse cultural backgrounds, which would enable more robust inferences about the impacts of particular programs, policies, or even pedagogical approaches on students' SEL development.

2. Purpose and Objectives

The CORE districts—a consortium of eight California school districts collectively serving more than one million students attending roughly 1,500 schools—provide a unique opportunity to examine whether specific school-wide or classroom practices are effective in helping support students' SEL. In 2013, the CORE districts secured a waiver from key requirements of No Child Left Behind that enabled them to develop a holistic school quality measurement system. This system includes surveys of students' SEL, as well as surveys of students', parents', and staff's perceptions on school culture and climate. CORE's SEL measures asks students in grades 3-12 to self-report their responses to 25 items on a 5-point Likert scale; the items cover four SEL domains: growth mindset, self-efficacy, self-management, and social awareness.

Because research shows that teachers play a central role in helping establish classroom and school environments that contribute to students' social and emotional development (Blazar & Kraft, 2017; Kraft, 2017; Jackson, 2014), this paper aims to leverage the CORE SEL survey in order to examine whether we can reliably estimate teacher effects on students' SEL. Specifically, we evaluate whether we can develop a teacher-level measure of student growth in SEL by applying statistical methods (i.e., value-added models) often used to assess teacher impacts in math and English language arts to students' self-reported surveys of SEL. Similar to existing work (Loeb, Christian, Hough, Meyer, Rice, & West, 2018), we also compare the variance components at the school and teacher levels in order to examine whether most of the differences in teachers' effects on students' SEL is due to the schools they teach in, or due to the individual teacher's impact. By assessing whether we can develop a sound approach for measuring teachers' impacts on students' SEL, we aim to contribute to the growing body of knowledge about appropriate and innovative uses of data on students' non-cognitive and social-emotional learning.

3. Data and Sample

We analyze data from the 2016-17 and 2017-18 administrations of CORE's student SEL survey, which include responses from more than 400,000 students each year. We focus on students in Grades 4 and 5, since these are typically self-contained classrooms with one teacher of record for a group of students. By first establishing whether we can estimate teacher-level effects of students' SEL for self-contained classrooms, we build a statistical foundation for estimating such effects before exploring possible generalizations and applications to more complex teacher-student links (e.g., in middle and high school, in co-teaching environments, or for special education teachers pushing into general education classrooms).

1. Research Design and Analysis

We estimate separate value-added models for each of the four SEL constructs assessed—growth mindset, self-efficacy, self-management, and social awareness. In order to assess the relative strength and robustness of the SEL value-added models, we compare the results of these models to value-added models for math and English language arts. Finally, we compare the results from these models to school-level value-added models of students' SEL from prior work (Loeb et al., 2018) in order to assess the degree to which differences in teachers' impacts on students' SEL is due to the individual teacher or to the school in which he or she teaches. The Appendix provides additional details on the model specification.

2. Results

Preliminary results indicate there are substantive differences across teachers in students' SEL growth, similar in magnitude to those for growth in academic achievement; across the four SEL constructs, we estimated a standard deviation of teacher effects in the range of 0.09 and 0.24 times the standard deviation of the level of SEL outcome measures across students (compared to standard deviations ranging between 0.11 and 0.18 for math and ELA). However, the goodness of fit of these value-added models is considerably lower for SEL than for measures of academic achievement.

3. Conclusions

As conversations around appropriate and valid uses of measures of students' SEL continue—particularly in response to the passage of The Every Student Succeeds Act (ESSA) in 2015—this paper aims to provide some preliminary evidence as to whether we can measure teachers' impact on students' SEL. In addition, these results will begin to shed light on whether embedding SEL into teachers' existing pedagogical practices might produce measurable impacts on students'

SEL, which further informs how schools and districts think about programs, policies, and initiatives to intervene on students' SEL.

References

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

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.

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

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

Jackson, C. K. (2014). *Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina* (NBER Working Paper No. 18624). Cambridge, MA: National Bureau of Economic Research.

Kraft, M. A. (2017). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*.

Loeb, S., Christian, M. C., Hough, H., Meyer, R. H., Rice, A. B., & West, M. R. (2018). *School effects on social-emotional learning: Findings from the first large-scale panel survey of students*. Stanford, CA: Policy Analysis for California Education.

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, 56,101-119.*

McKown, C. & Taylor, J. J. (2018). Introduction to the special issue on social-emotional assessment to guide educational practice. *Journal of Applied Developmental Psychology, 55*, 1-3.

Appendix

We model the impacts of teachers on SEL and academic outcomes using the following value-added regression model:

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

where student *i* is associated with teacher *j* in year *t*; y_{cijt} is the outcome in construct *c* for student *i* associated with teacher *j* in year *t*; y_{Cijt-1} is a 1 × 6 vector of outcomes for student *i* in year *t*-1 in all four SEL constructs and two academic subjects (mathematics and English Language Arts); X_{ijt} is a vector of characteristics of student *i* in year t; α_{cjt} is a fixed effect for teacher *j* for 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).

This equation is estimated using an 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. This is to avoid attenuation bias from measurement error in the measures of lagged SEL and academic outcomes included on the right-hand-side of the estimated regression. We center the teacher 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 teacher in the regression sample. As a result, the teacher fixed effects are measured relative to the average teacher effect across the teachers in the sample. We use these centered teacher effect estimates as the measures of teacher growth for each of the six SEL and academic outcomes.

Paper 3
Is Social-emotional Learning (SEL) a State or a Trait?
Examining the Stability of SEL across Three Years

Megan Kuhfeld, NWEA
James Soland, NWEA

1. Background and Context

Considerable attention and resources are being devoted to social-emotional learning (SEL) by educational practitioners and policymakers. Research already shows that, for many SEL constructs like growth mindset and academic belonging, targeted interventions can shift mean scores positively pre- and post-intervention (Yeager & Walton, 2011). Our study asks a different question, namely whether student rank orderings based on SEL scores are stable over time (trait-like) versus sensitive to situational and environmental factors (state-like). The state-trait balance has important implications for how to foster SEL, as well as evaluate the effect of teachers and schools on these constructs. Theoretically, the more trait-like a construct, the more likely the effects of an intervention are to persist, but the less likely students are to be reordered on that construct over time, which has implications for equity and score gaps. Conversely, the more state-based a construct, the more students can be reordered over time and the more scores at a given time point are due to situational and environmental factors. These situational factors are oftentimes more within the control of teachers and schools to change (e.g., classroom environment and day-to-day personal interactions with students). For instance, if an SEL constructs is more state-like, then devoting a month to improving students' self-management may be less impactful than changing classroom and school environmental factors that support self-management, even if the gains over the month are non-trivial.

2. Purpose

The purpose of this study is to quantify the degree to which four SEL competencies used in the California Office to Reform Education's (CORE) accountability system are state-like versus trait-like. In so doing, we examined two research questions. First, what proportion of the variance in each SEL construct is explained by a stable trait-like factor, and how do those proportions differ across constructs? Second, how does the relative stability of the SEL constructs compare with that of math and reading achievement?

3. Data and Sample

We used three waves (2014-15, 2015-16, 2016-17) of data collected from 33,534 total students in 3rd to 11th grade from Santa Ana Unified School District (SAUSD), a member of the CORE consortium. The district is urban, high-poverty, and serves a high proportion of English learners. There are 54 schools in our sample. Table 1 presents demographic information on students in the sample.

We used survey responses from a yearly SEL survey measuring self-efficacy, growth mindset, self-management, and social awareness. Measurement properties of the survey scales are provided in Table 2. In addition to using SEL surveys, the district administers the Measures of Academic Progress (MAP) Growth assessment, a vertically scaled, computer-adaptive test of reading and mathematics achievement.

4, Research Design and Analysis

We used a latent-state trait (LST) structural equation modeling approach, which has been demonstrated with a wide range of psychological and educational constructs, including personality traits (Prenoveau, Craske, Zinbarg, Mineka, Rose, & Griffith, 2011), self-esteem (Prenoveau, 2016), and math ability (Bailey, Watts, Littlefield, & Geary, 2014). We examined used the trait-state-occasion (TSO; Cole, Martin, & Steiger, 2005) parameterization presented in Prenoveau (2016), which is shown in Figure 1. In the TSO model, the relative standing S_{k} can be perfectly explained by knowing (a) standing on the SEL construct trait factor (T), and (b) standing on the time point's occasion factor, O_{k}. Therefore, the variance of the state factor can be partitioned into the variance explained by the state factor and the variance not explained by the state factor (e.g., occasion variance). To allow for the possibility that occasion factors may still be correlated after accounting for the trait factor standing, autoregressive paths are estimated between occasion factors at adjacent timepoints. Therefore, after the first time point, the variance of the state factor is partitioned into three components: (a) variance explained by the trait, (b) variance explained by the previous occasion (situational effects that carry over), and (c) unexplained variance specific to the time point.

We estimated the TSO models for all four SEL constructs as well as math and reading achievement. We calculated proportion of variance attributable to trait, occasion, and prior state factors.

5. Results

Table 3 presents the proportion of state variance explained by trait, occasion, and prior state. These results are presented for the four SEL constructs of interest and academic achievement. At time 1, the proportion of SEL state variance attributable to the trait factor ranged from 36% (social awareness) to 51% (growth mindset). In general, a higher proportion of the state variance is attributable to the trait factor for achievement test scores than for SEL constructs. Roughly 70-85% of the variance in achievement test scores is attributable to the trait factor, which implies that the rank-order stability of academic achievement is high, and the impact of situational influences on the rank-ordering within a time point is low. In other words, while the vast majority of the variation across time in achievement is due to person-characteristics (trait), approximately half of the variation in SEL constructs is due to situational and environmental factors around the time of measurement.

6. Conclusion

Our study has demonstrated that there is balance between state and trait components of SEL-based rank orderings over time. The presence of state and trait components suggests that schools wishing to produce sustained improvements in SEL scores would likely need to find a balance between early SEL interventions that impact the trait latent variable (and may be less likely to fade out) with episodic supports for students that improve the contextual factors influencing SEL. Even in the absence of intervention, our results suggest that educators could benefit from understanding and focusing on the environmental factors that contribute to a student's SEL-based rank.

References

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychological Science*, *25*(11), 2017–2026.

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait–state models: Introducing a trait–state–occasion model. *Psychological Methods, 10*, 3–20. doi:10.1037/1082-989X.10.1.3

Prenoveau, J. M., Craske, M. G., Zinbarg, R. E., Mineka, S., Rose, R. D., & Griffith, J. W. (2011). Are anxiety and depression just as stable as personality during late adolescence? Results from a three-year longitudinal latent variable study. *Journal of Abnormal Psychology*, *120*(4), 832.

Prenoveau, J. M. (2016). Specifying and interpreting latent state–trait models with autoregression: An illustration. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 731–749.

Yeager, D. S., & Walton, G. M. (2011). Social-Psychological Interventions in Education They're Not Magic. *Review of Educational Research, 81*(2), 267–301.

Table 1
Statistics on Analytic Sample

	2014-15	2015-16	2016-17
Grade range	4 to 10	3 to 11	3 to 11
Prop. students with growth mindset scores	13,173	24,210	22,659
Prop. students with self-efficacy scores	13,168	24,200	22,656
Prop. students with self-management scores	13,338	24,232	22,697
Prop. students with social awareness scores	13,190	24,206	22,657
Prop. Asian	0.020	0.019	0.022
Prop. Black	0.002	0.002	0.002
Prop. Hispanic	0.966	0.965	0.963
Prop. White	0.006	0.007	0.007
Prop. Female	0.501	0.503	0.505
Math percentile (median)	29	31	33
Reading percentile (median)	29	25	29

Table 2

Statistics on SEL Survey Used by the District

	Growth Mindset			Self-efficacy			Self-management			Social Awareness		
	2014-15	2015-16	2016-17	2014-15	2015-16	2016-17	2014-15	2015-16	2016-17	2014-15	2015-16	2016-17
Reliability (Cronbach)	0.65	0.65	0.68	0.86	0.85	0.85	0.84	0.83	0.83	0.80	0.79	0.80
Sum Scores												
Mean	3.51	3.50	3.57	3.43	3.40	3.39	3.97	3.91	3.92	3.68	3.68	3.70
Standard deviation	0.89	0.90	0.92	0.92	0.93	0.93	0.67	0.68	0.68	0.63	0.65	0.67
Skew	-0.21	-0.21	-0.31	-0.26	-0.22	-0.22	-0.85	-0.65	-0.68	-0.55	-0.46	-0.54
Kurtosis	2.52	2.48	2.52	2.49	2.42	2.46	3.96	3.20	3.38	3.73	3.43	3.60
First eigenvalue	2.14	2.13	2.28	2.98	2.91	2.97	4.58	4.37	4.51	3.73	3.60	3.82
Second eigenvalue	0.81	0.80	0.74	0.40	0.42	0.40	0.88	0.88	0.88	0.86	0.85	0.82

Table 3

Proportion of Variance in SEL Constructs Explained by State Versus Trait Factors

	Growth Mindset	Self-Efficacy	Self-Management	Social-Awareness	Math Achievement	Reading Achievement
First State/Time Period						
Trait	0.513	0.464	0.382	0.355	0.731	0.837
Occasion	0.487	0.536	0.618	0.645	0.268	0.162
Second State/Time Period						
Trait	0.513	0.464	0.382	0.355	0.774	0.837
Occasion	0.478	0.510	0.527	0.587	0.204	0.156
Prior State	0.009	0.026	0.091	0.058	0.022	0.007
Third State/Time Period						
Trait	0.513	0.464	0.382	0.355	0.841	0.808
Occasion	0.485	0.525	0.547	0.611	0.159	0.170
Prior State	0.002	0.011	0.071	0.034	0.001	0.021

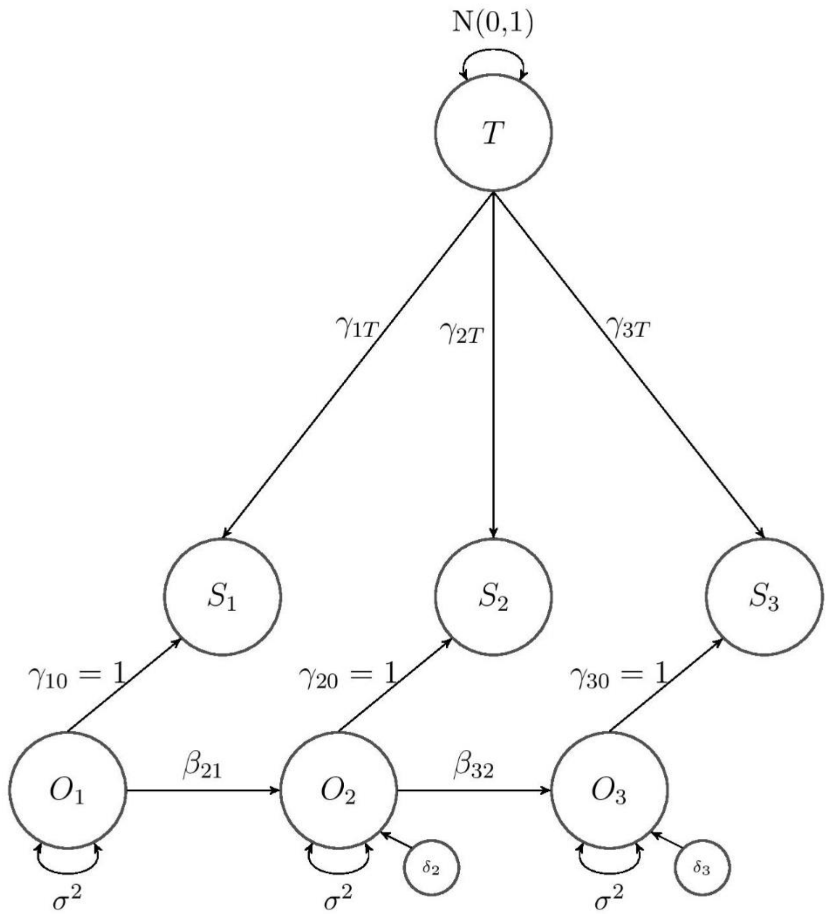

Figure 1. Trait–state–occasion (TSO) model. k = time point (1, 2, and 3); S_k = latent state SEL at time point k; O_k = latent occasion SEL at time point k; T=latent trait SEL; γ = regression coefficients for pathways from T to S_k (γ_{kT}), and O_k to S_k (γ_{kO}); $\beta_{k,k-1}$ = regression coefficient for the autoregressive path from O_k to O_{k-1}; and δ_k = disturbance term for O_k. Each state latent variable (S_k) is constrained to have 0 residual variance such that the state variance at a given time point is entirely determined by the variance of trait, occasion, and prior state latent variables. Constraints are included so that $\gamma_{1T} = \gamma_{2T} = \gamma_{3T}$ and the total residual variance (σ^2) is equal across the three time points. Each of the state factors has a measurement model that is not shown here to avoid cluttering the path diagram.

Paper 4
Do Math Self-efficacy and Achievement Develop in Tandem?
Evidence and Implications

James Soland, NWEA

1. Background and Context

Initiatives to develop students' SEL are proliferating in education policy, practice, and research. Among constructs related to SEL, academic self-efficacy has gained prominence due to its association with long-term educational achievement and attainment (Alivernini & Lucidi, 2011; Farrington et al., 2012; Kitsantas, Cheema, & Ware, 2011). For instance, students who struggle academically often have lower self-efficacy, defined as the confidence they feel in their ability to successfully complete academic tasks (Bandura, 1993; Meyer, Wang, & Rice, 2018). Given that self-efficacy has been shown to differ by subgroups, there is some evidence that achievement gaps are partially a function of diminished self-efficacy, including between male and female students in mathematics (Li, 1999). Based in part on this research, measures of self-efficacy have also gained traction in current district efforts to develop student SEL. For example, districts participating in the California Office to Reform Education (CORE) include self-efficacy scores in their accountability policies, which were developed through a waiver of the No Child Left Behind Act of 2001 (West, Buckley, Krachman, & Bookman, 2017).

2. Purpose and Research Questions

Despite the importance of self-efficacy, we do not fully understand how students' baseline levels of self-efficacy influence their growth in math achievement over time (and vice-versa), nor whether gains in self-efficacy over time are associated with gains in achievement (and vice-versa). In particular, we do not have much evidence on whether self-efficacy affects test performance or the reverse. Does a low score on a math test appear to negatively impact a student's self-efficacy at a later time point, or is lower self-efficacy associated with decreases in math achievement at later time points? Even less is known about whether these developmental processes differ by school. We begin to close that gap by asking three research questions:

1. Is there evidence that self-efficacy is measured consistently across grades?
2. How does self-efficacy change across middle school?
3. How do changes in self-efficacy affect growth in math achievement and vice-versa?
4. Does the relationship between self-efficacy and math achievement differ by school?

3. Data and Sample

We answer these questions using SEL survey data collected over four years in Santa Ana Unified School District, which is a member of the CORE consortium. The district is high-poverty and serves a high proportion of English learners. Table 1 presents information on demographics and Table 2 on the SEL measures. Each student also took the Measures of Academic Progress (MAP) Growth, a computer-adaptive vertically scaled achievement test in math. For the purposes of our study, we examined SEL and achievement trends for cohorts of students beginning in 6th and 7th grade.

4. Research Design and Analysis

Establishing that the survey items measure the construct of self-efficacy consistently across ages is an important precursor to any longitudinal modeling. To answer Question 1, we conducted longitudinal measurement invariance testing on our self-efficacy scores. This model included a measurement model with each of the four observed self-efficacy item responses loading on a latent self-efficacy variable separately by time period. We then imposed constraints related to configural (factor structure), weak (loadings), and strong (thresholds) invariance, then compared model fit using fit statistics common in the literature for nested models (Bentler, 1990; Hu & Bentler, 1998; Koh & Zumbo, 2008).

For Question 3, we compared the fit of the latent curve models (LCMs) detailed in path diagrams in Figures 1-3 (Curran et al., 2014). The first is an LCM with a growth model for achievement and treating self-efficacy scores as time-varying covariates (we also fit this model, but with self-efficacy as the outcome and math score as the covariate). The second is a bivariate unconditional latent curve model that models a separate growth curve for self-efficacy scores and math achievement, giving each construct its own developmental submodel (Curran et al., 2014). The third model is a bivariate unconditional autoregressive latent trajectory model, which adds paths from lagged self-efficacy to current math score and vice-versa. Model 3 therefore allows us to constrain paths from lagged self-efficacy to current math score (and vice-versa) to zero systematically and, thereby, gain evidence on whether low self-efficacy leads to lower math achievement or the reverse (Curran et al., 2014). Finally, we fit the same three models, but using a multilevel LCM to examine the variance between and within schools.

5. Results

Table 3 presents the results of the measurement invariance testing. While there is no significant decrement in fit from the configural to the weak model, there is a decrease from weak to strong. However, a model that imposes partial measurement invariance by loosening restrictions on thresholds for three item-by-time combinations largely addresses the issue.

Turning to results for Question 2, Figure 1 presents plots of self-efficacy over time by grade. These figures suggest there is a clear downward trend to self-efficacy as students get older. Table 3 presents results from measurement invariance tests.

Table 4 presents correlations among parameters from the model in Figure 2, which indicate that students with high starting math achievement also have high starting self-efficacy scores. Further, the higher a student's initial self-efficacy, the higher the slope on growth in math achievement. While we are awaiting a fourth year of achievement test data to complete the model in Figure 3, initial results using three years of data suggest a model fits better that assumes low self-efficacy is associated with low achievement at a later time point, not vice-versa. Preliminary multilevel LCMs indicate little of the variance in self-efficacy growth trajectories is between schools.

6. Conclusion

Initial results indicate that self-efficacy scores show a clear downward trend over time. Models confirm that these scores are suited to latent growth curve modeling. More importantly, self-efficacy and achievement appear to develop in relation to each other, and early findings suggest self-efficacy may impact math achievement rather than the reverse. Such results can shed light on how growth processes in SEL and achievement develop over time together, including providing insight into how to intervene to improve achievement, self-efficacy, or both.

References

Alivernini, F., & Lucidi, F. (2011). Relationship between social context, self-efficacy, motivation, academic achievement, and intention to drop out of high school: A longitudinal study. *The Journal of Educational Research*, *104*(4), 241–252.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117–148.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238.

Curran, P. J., Howard, A. L., Bainter, S. A., Lane, S. T., & McGinley, J. S. (2014). The separation of between-person and within-person components of individual change over time: A latent curve model with structured residuals. *Journal of Consulting and Clinical Psychology*, *82*(5), 879.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance–A Critical Literature Review*. ERIC.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424.

Kitsantas, A., Cheema, J., & Ware, H. W. (2011). Mathematics achievement: The role of homework and self-efficacy beliefs. *Journal of Advanced Academics*, *22*(2), 310–339.

Koh, K. H., & Zumbo, B. D. (2008). Multi-group confirmatory factor analysis for testing measurement invariance in mixed item format data. *Journal of Modern Applied Statistical Methods*, *7*(2), 12.

Li, Q. (1999). Teachers' beliefs and gender differences in mathematics: A review. *Educational Research*, *41*(1), 63–76.

Meyer, R. H., Wang, 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*.

West, M. R., Buckley, K., Krachman, S. B., & Bookman, N. (2017). Development and implementation of student social-emotional surveys in the CORE Districts. *Journal of Applied Developmental Psychology*.

Table 1

Statistics on Analytic Sample

	2014-15	2015-16	2016-17	2017-18
Grade range	4 to 10	3 to 11	3 to 11	3 to 11
Number of students with self-efficacy scores	7,508	10,367	8,418	9,044
Prop. Male	0.487	0.492	0.500	0.478
Prop. ELL	0.297	0.289	0.227	0.202
Prop. Special Ed.	0.094	0.110	0.102	0.066
Prop. Hispanic	0.959	0.969	0.957	0.965
Math percentile (median)	29	31	33	30
Reading percentile (median)	29	25	29	27

Table 2

Statistics on SEL Survey Used by the District

	Self-efficacy			
	2014-15	2015-16	2016-17	2017-18
Reliability (Cronbach)	0.86	0.85	0.85	0.87
Sum Scores				
Mean	3.525	3.347	3.294	3.223
Standard deviation	0.939	0.948	0.935	0.915
Skew	-0.379	-0.192	-0.146	-0.098
Kurtosis	2.496	2.423	2.527	2.662
First eigenvalue	2.98	2.91	2.97	2.99
Second eigenvalue	0.40	0.42	0.40	0.31

Table 3

Measurement Invariance Results

Model	N	Test of Overall Fit			Likelihood Ratio Test			Fit Indices		
		χ^2	d.f.	p	$\Delta\chi^2$	d.f.	p	RMSEA	CFI	TLI
M1. Configural	4157	261.013	74	0.000				0.025	0.997	0.995
M2. Weak	4157	250.775	83	0.000	13.416	9	0.145	0.022	0.997	0.996
M3. Strong	4157	509.441	128	0.000	262.607	45	0.000	0.027	0.994	0.994
M4. Partial Strong										
Compared to Strong (M3 nested in M4)	4157	296.922	114	0.000	220.340	12	0.000	0.020	0.997	0.997
Compared to Weak (M4 nested in M2)	4157	296.922	114	0.000	49.583	31	0.018	0.020	0.997	0.997

Table 4

Initial Results from the Model in Figure 2

Parameter	Standardized Coefficient	
Intercept math with slope math	0.229	*
Intercept self-efficacy with slope self-efficacy	-0.284	*
Intercept math with intercept self-efficacy	0.435	*
Slope self-efficacy with intercept math	-0.032	
Slope math with intercept self-efficacy	0.204	*
Slope math with slope self-efficacy	-0.099	

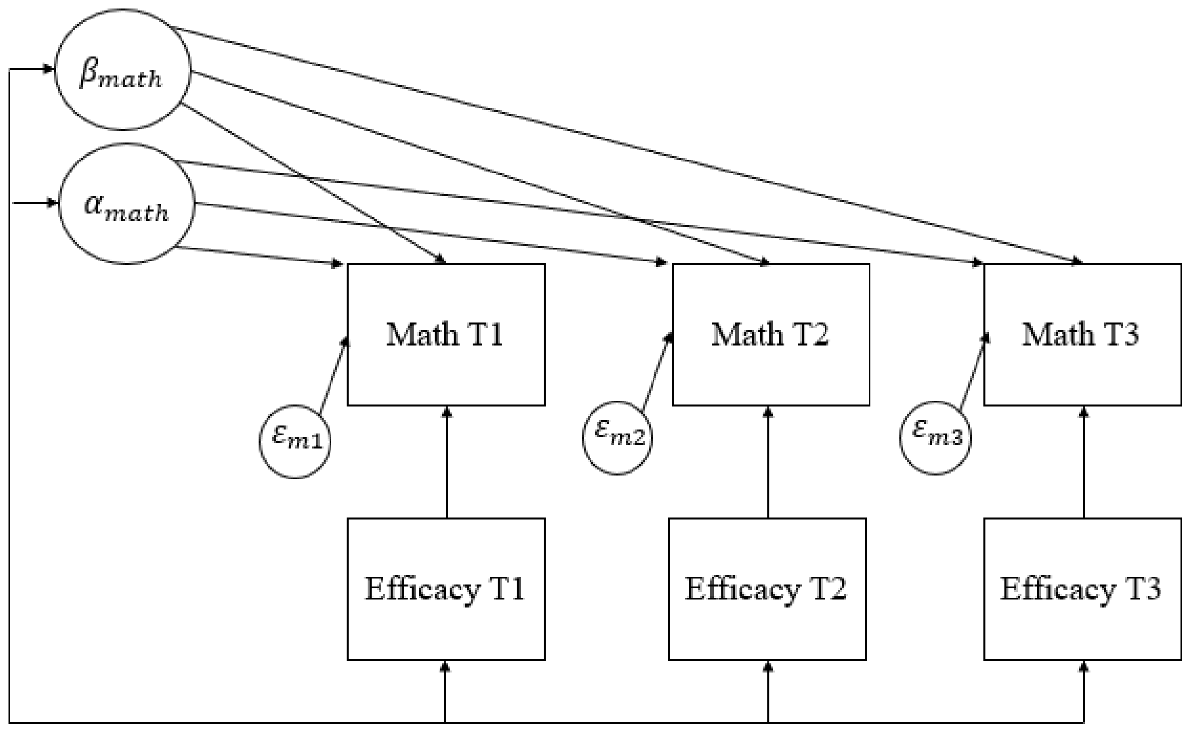

Figure 1. Path diagram for the time-varying covariates model.

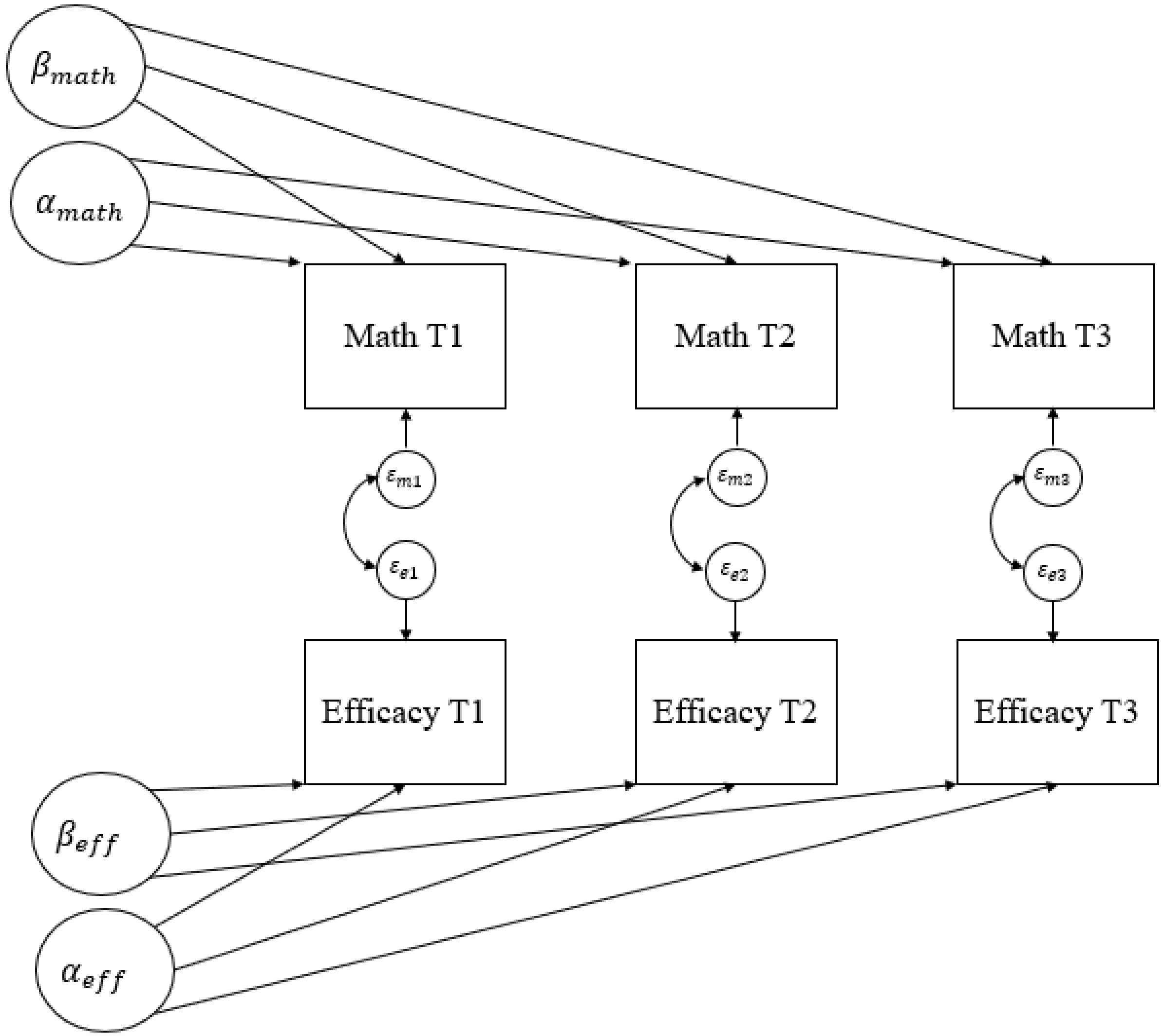

Figure 2. *Path diagram for the bivariate unconditional latent curve model.*

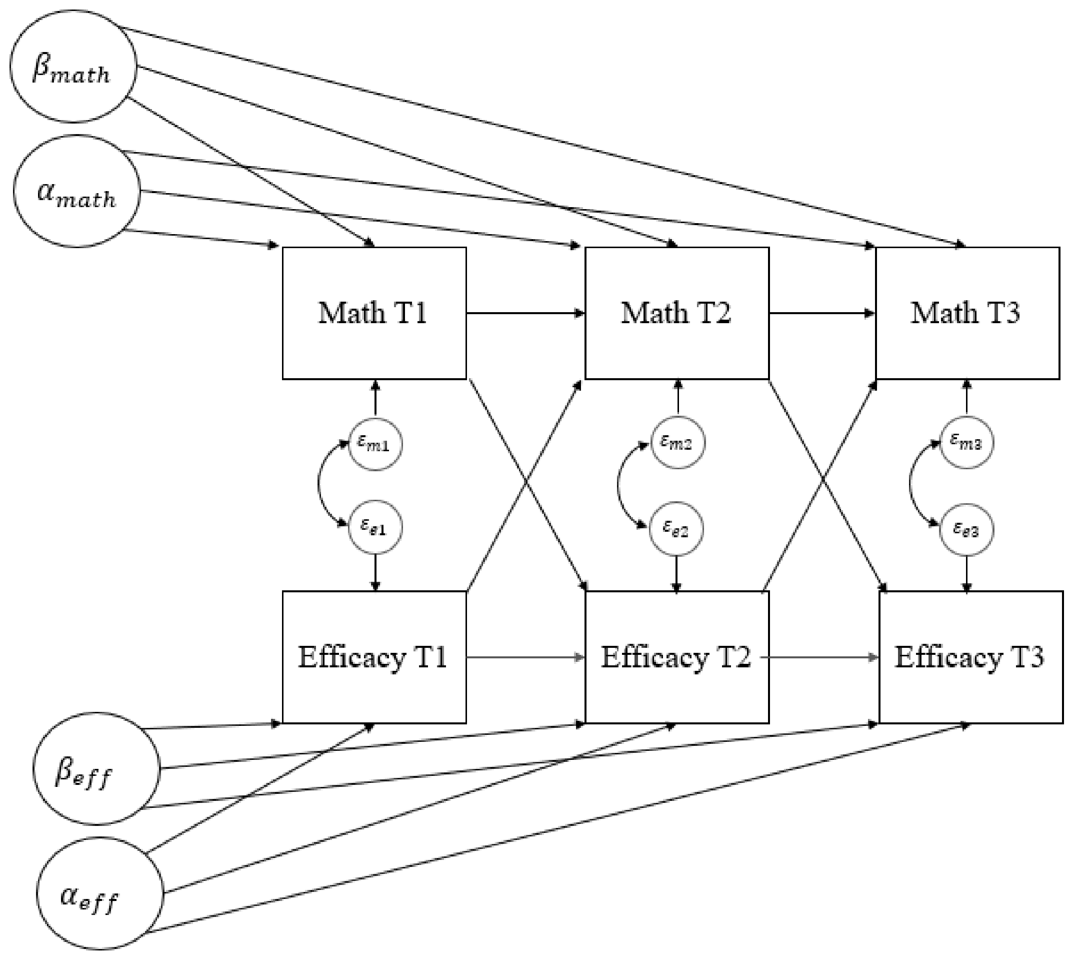

Figure 3. Path diagram for the bivariate unconditional autoregressive latent trajectory model.

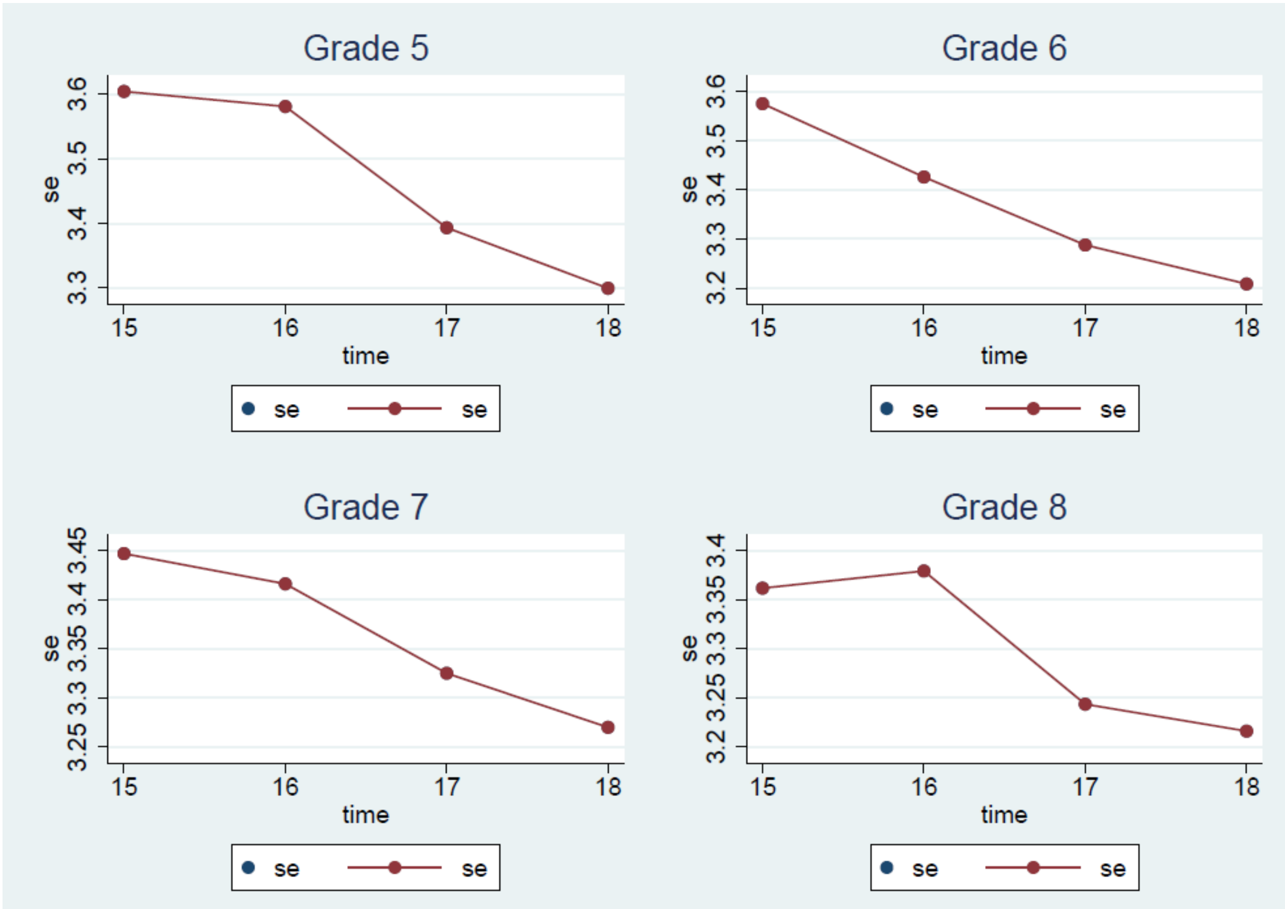

Figure 4. Trends in self-efficacy by grade cohort.

Note. Grade refers to the student's grade level in 2015 (first time point).