

English Language Learners, Self-efficacy, and the Achievement Gap:

Understanding the Relationship between Academic and Social-emotional Growth

By Jim Soland

The Collaborative for Student Growth at NWEA Working Paper series is intended to widely disseminate and make easily accessible the results of researchers' latest findings. The working papers in this series have not undergone peer review or been edited by NWEA. The working papers are widely available, to encourage discussion and input from the research community before publication in a formal, peer-reviewed journal. Unless otherwise indicated, working papers can be cited without permission of the author so long as the source is clearly referred to as a Collaborative for Student Growth at NWEA working paper.

Soland, J. (2019). *English Language Learners, Self-efficacy, and the Achievement Gap: Understanding the Relationship between Academic and Social-emotional Growth* (The Collaborative for Student Growth at NWEA Working Paper).

Abstract

Due in part to the challenges associated with learning a new language, English language learners (ELLs) typically begin school with lower achievement than their non-ELL peers, and those achievement gaps often close slowly if at all. A separate body of research shows that achievement is associated with social-emotional learning constructs like self-efficacy, yet this relationship has rarely been examined for ELLs. In this study, multivariate models that jointly estimate growth in achievement and self-efficacy during middle school are used to see how underlying developmental processes relate for ELLs. Results indicate that self-efficacy tends to decline for all students despite growth in math and reading, and that achievement and self-efficacy are much lower for ELLs. Further, there is evidence that slower growth in math and reading for ELLs is associated with their low self-efficacy at the beginning of middle school (self-efficacy mediates the association between ELL status and achievement growth). Implications for closing achievement gaps between ELLs and non-ELLs are discussed.

Keywords: English language learners, achievement gaps, student growth, self-efficacy, social-emotional learning (SEL)

Running Head: ELL Self-efficacy

English Language Learners, Self-efficacy, and the Achievement Gap:

Understanding the Relationship between Academic and Social-emotional Growth

English language learners (ELLs) are a rapidly growing student subgroup in United States schools (Cook, Linqunti, Chinen, & Jung, 2012; Linqunti & Cook, 2013). These students face the daunting challenge of learning a new language in tandem with developing academic proficiency in core subjects like reading and mathematics (Cook, Boals, & Lundberg, 2011). Given these challenges, ELLs often have lower mean achievement test scores in reading and mathematics (Aguirre-Muñoz & Boscardin, 2008; Cook et al., 2011). Partially as a result, ELLs also tend to have higher dropout rates and lower college attendance (Callahan, 2013; Kanno & Cromley, 2013). Research further suggests that results on these long-term outcomes are partially due to school contextual factors like access to appropriate course material (Aguirre-Muñoz & Boscardin, 2008; Callahan, 2013). With such barriers to educational achievement and attainment, there is a strong need for evidence on ways that growth in math and reading for ELLs can be accelerated such that achievement gaps with non-ELL peers close over time.

Studies also show that academic self-efficacy—defined as a students’ confidence in their ability to attain a certain educational goal or outcome—is strongly associated with achievement (Kitsantas, Cheema, & Ware, 2011). This construct has therefore been highlighted in much of the emerging policy and practice work around developing students’ social-emotional learning (SEL) skills (Farrington et al., 2012; Soland, Hamilton, & Stecher, 2013; Soland, 2018; West, Buckley, Krachman, & Bookman, 2017). Further, some studies show that higher self-efficacy is associated with faster growth in math and reading (Caprara et al., 2008). Thus, improving self-efficacy may be useful in helping close achievement gaps over time.

Running Head: ELL Self-efficacy

Despite the importance of academic growth for ELLs, few studies consider the relationship between achievement and self-efficacy for this subgroup. This study begins to close the gap in the literature by using four years of vertically scaled achievement test data and self-efficacy survey scores to examine how the two outcomes are related, including how they grow in tandem for ELLs. The data are for ELLs attending middle school in a large urban district in California where more than 90% of students are Hispanic and low-income, and over a third of those students enter Kindergarten as ELLs. In short, while the sample may not be representative of the average U.S. district, it represents exactly the sort of context in which the issues investigated are of paramount importance. Using these data, three research questions are addressed:

1. Does self-efficacy differ for ELLs versus non-ELLs?
2. Do growth patterns in self-efficacy and achievement differ for ELLs versus non-ELLs?
3. Does self-efficacy mediate growth in achievement for ELLs?

In the remainder of the study, background on what is known about achievement, self-efficacy, and their relationship for ELLs is discussed, then methods and findings are presented followed by a discussion of their implications.

Background

Academic self-efficacy is one of the most studied SEL constructs. Therefore, only research directly pertinent to the association between achievement and self-efficacy for ELLs will be reviewed. Particular focus is given to the ways that race/ethnicity relate to self-efficacy and achievement, as well as how self-efficacy can mediate the relationship between race and achievement.

ELL Achievement

Running Head: ELL Self-efficacy

On average, ELLs perform below grade level in every subject tested for federal accountability (Cook et al., 2011). On the National Assessment of Educational Progress (NAEP), nearly half of ELLs in 4th grade and over 70% in 8th grade performed “below basic” in math (Cook et al., 2011; Fry, 2008). Gaps between ELLs and non-ELLs in reading were also large, with ELLs scoring 36 points below their native-speaking peers in 4th-grade reading and 44 points below in 8th grade reading on the 2011 NAEP. The vast majority of ELLs are below proficient based on state reading achievement tests, though proficiency rates increase substantially as ELLs become proficient on English language tests (Cook et al., 2011; Fry, 2008). Given the state of ELL achievement, it is not entirely surprising that ELLs are twice as likely to drop out as their native English-speaking peers (Rumberger, 2006).

One complexity associated with comparing achievement between ELLs and non-ELLs is that, as ELLs learn English and catch up to their peers in reading, they often exit ELL status altogether, becoming what is known under federal law as “Reclassified Fully English Proficient” or R-FEP (Umansky & Reardon, 2014). Thus, under the definition of ELL used in federal law, there will always be an achievement gap between ELLs and non-ELLs because, once the gap in reading is closed, most ELLs have reclassified. As a result, many studies examine achievement, and changes in relative achievement over time (gaps), using students who were ever deemed to be ELL, which makes the group more stable (Loeb, Soland, & Fox, 2014; Matta & Soland, 2018). That is, contrasts in achievement between ELLs and non-ELLs involve defining the former as a student who was ever designated as an ELL, including after the student has been reclassified as R-FEP. Generally, the shifting nature of the ELL subgroup indicates both the complexities inherent in estimating achievement gaps for this subgroup and the importance of

Running Head: ELL Self-efficacy

closing those gaps rapidly in order to reclassify students such that they have the language skills to succeed in mainstream classrooms.

Self-efficacy, Achievement, and Race/Biological Sex

Self-efficacy refers to how individuals judge their own abilities to perform certain tasks or actions (Bandura, 1986, 1993). The higher a person's self-efficacy, the more that person believes he or she will be able to successfully complete a certain action or perform at a particular level (Bandura, 1993). Bandura (1993, 1994) argued that self-efficacy is the foundation of human motivation: without belief in one's ability to accomplish a task, there is little incentive to undertake it. In education, the construct of self-efficacy measures a student's confidence in his or her ability to attain a certain educational goal or outcome, such as the ability to do well on a test or earn good grades in class. As an example of this theory in action, Zimmerman (2000) showed that student self-efficacy predicts motivation to learn, including students' activity choices, effort, persistence, and emotional reactions to difficult situations. Given the impact of self-efficacy on motivation and persistence, the construct is associated with grades, educational attainment, and achievement (Zimmerman, Bandura, & Martinez-Pons, 1992). Research further shows that growth in math and reading over time are associated with growth in self-efficacy as students move through school, suggesting that improving self-efficacy may be useful in closing achievement gaps (Caprara et al., 2008).

Research also finds that academic self-efficacy is associated with race and ethnicity, especially in particular educational contexts. For example, Vuong, Brown-Welty, and Tracz (2010) found that demographic factors influence self-efficacy, and that self-efficacy beliefs affect the GPA and persistence rates of first-generation college students. These persistence rates were further impacted by school contextual factors like size of the institution (Vuong, Brown-

Running Head: ELL Self-efficacy

Welty, & Tracz., 2010). Kitsantas, Cheema, and Ware (2011) showed that black students often report lower self-efficacy than white students, and that introducing self-efficacy concepts into math assignments helped close the achievement gap in their sample. This finding provides evidence that educator interactions with students can influence self-efficacy, suggesting teachers may be in a position to improve self-efficacy among minority students (Kitsantas et al., 2011; Schunk & Meece, 2006).

Similar findings on race and biological sex have emerged when specifically studying middle school students. Research shows that SEL constructs, including self-efficacy, can begin to shift as students enter middle school. For instance, during the first year of middle school, students in one study described mathematics as less valuable and reported lower levels of effort in the subject (Pajares & Graham, 1999). In the same study, high-achieving students reported lower levels of self-efficacy during that time span. Research also shows that the relative math and science self-efficacy of boys versus girls can shift during middle school, though the direction is not always consistent (Britner & Pajares, 2006). Related work indicated that, in middle school, white students had higher science self-efficacy than their non-white peers (Britner & Pajares, 2001). Oftentimes, these differences in self-efficacy on the basis of race and biological sex are evident when students enter middle school (Usher & Pajares, 2006).

There is also evidence that academic self-efficacy can mediate the association between achievement and other predictors of achievement, including race and biological sex. For instance, Pajares & Miller (1994) showed that biological sex and prior experience influenced mathematics problem solving largely through the mediational role of self-efficacy. Related work indicated that mathematics self-efficacy is a mediator between mathematics attitude and achievement with potential differences in relevant parameter estimates based on student

Running Head: ELL Self-efficacy

demographics (Randhawa, Beamer, & Lundberg, 1993). Similarly, Fast et al. (2010) showed that the effect of classroom environment on achievement was mediated by academic self-efficacy.

ELLs and Factors Related to Self-efficacy

Only one prominent study to date has examined the relationship between self-efficacy and achievement for ELLs. Lewis et al. (2012) found that having caring teachers bolstered self-efficacy in math for ELLs, which in turn shared a positive association with math test scores. The same study also showed that teacher caring's association with achievement was mediated less by self-efficacy for ELLs than non-ELLs.

Although there are not many studies that directly examine the academic self-efficacy of ELLs, a range of studies provide indirect evidence that self-efficacy could be different (and, most plausibly, lower) for ELLs. First, there is an established association for ELLs between SEL constructs and educational outcomes like achievement and reclassification to R-FEP status. For example, studies show that SEL mindsets like academic belonging are factors in how long it takes students to attain English proficiency and whether they finish high school (Castro-Olivo, Preciado, Sanford, & Perry, 2011). Other research has found that self-efficacy can shift as students engage in learning a foreign language and develop mastery over time (Mercer, 2012). Perhaps most relevant to the current study, Niehaus and Adelson (2013) used structural equation models (SEMs) to examine the relationship between achievement, academic self-concept (a construct directly related to self-efficacy), and the presence of "social-emotional problems" (p. 810) among ELLs. Using the Early Childhood Longitudinal Study, Kindergarten Cohort (ECLS-K), they found a strong association between social-emotional problems and achievement, but a nonsignificant correlation between self-concept and achievement for ELLs.

Running Head: ELL Self-efficacy

There is also evidence that being considered an ELL can, at times, be stigmatizing for students, which could affect self-efficacy. Dabach (2014) conducted a qualitative study and found that placing ELLs in separate classrooms designed to facilitate their English language proficiency actually led to those students doubting their own intelligence. In similar work, Gandara and Orfield (2010) showed that the excessive segregation of ELLs in Arizona schools was likely harmful to students' achievement and social-emotional development. Relevant to the sample used in this study, research suggests that, especially in middle school when students become long-term ELLs, emotional needs are consistently denied or ignored (Valenzuela, 1999) and students begin to "face reality" when it comes to their status as ELLs (Rubinstein-Avila, 2003, p. 133). More generally, the tracking of ELLs in middle and high school has been shown to limit students' opportunity to learn, which could be affect academic self-efficacy (Callahan, 2005).

Data, Measures, and Analytic Strategy

Data

The data come from a large urban school district in California. In the district, more than 95% of the students are Hispanic, and more than 30% enter the district as ELLs. The study focuses on a cohort of students who began in 5th grade in the 2014-15 school year and ended the study in 8th grade in the 2017-18 school year. The cohort was not intact: students could move in and out of the sample at any time. A single grade cohort was used to examine growth in achievement and self-efficacy over time in order to avoid confounding across-grade differences in SEL scores (West et al., 2017). A middle school cohort in particular was utilized because the SEL survey administered by the district (described in more detail below) differed slightly for elementary and middle school grades; therefore, middle school was emphasized to ensure the

Running Head: ELL Self-efficacy

measure did not change between years. However, results were not sensitive to using a different cohort, nor to combining two cohorts. Table 1 presents descriptive statistics for the sample.

As discussed in the background section, a challenging with comparing relative achievement for ELLs versus non-ELL is that, as gaps close, ELLs often become reclassified as R-FEP. To address the issue, an approach used in prior research was taken, namely examining achievement contrasts between non-ELLs and students who were designated ELL at any point during the sample (Loeb et al., 2014; Matta & Soland, 2018). To make this approach more concrete, Table 2 shows a breakdown of the sample based on the ELL and non-ELL designation used. As the table demonstrates, non-ELLs consisted of students who were never considered ELL (English only, or EO), or who took a language screener upon school entry and reported speaking another language, but were not classified as ELL on the basis of a language proficiency test (Initially Fluent English Proficient, or I-FEP). Meanwhile, the ever-ELL category consists only of students who began 5th grade as ELLs (Limited English Proficient, or LEP). Over time, some of those students reclassified as R-FEP. As detailed below, sensitivity analyses were conducted by including and excluding R-FEP students from analyses.

Measures

Achievement. Students in the district took MAP Growth, an assessment of math and reading achievement. The tests are vertically scaled, which helps support the use of scores in growth models. MAP Growth is often administered in fall and spring terms, though this study mainly used spring scores since those tests were administered during the same time period as the self-efficacy survey. MAP Growth is a computer-adaptive test, which means students receive content matched to their estimated achievement level, helping avoid instances where students are receiving content that is extremely difficult or easy for them. In tandem, these attributes of MAP

Running Head: ELL Self-efficacy

Growth mean that growth can be estimated on a consistent and comparable scale for all time periods and grades in the study, even for students performing below grade level.

Estimated reliability of MAP Growth achievement test scores is high due to the computer-adaptive design. Each test takes approximately 40 to 60 minutes depending on the grade and subject area. Students respond to assessment items in order (without the ability to return to previous items), and a test event is finished when a student completes all the test items (typically 40 items for reading). Thus, the test is long enough that, when combined with its adaptivity, the student-specific standard error of measurement is generally small (Thum & Hauser, 2015). Because the test uses Item Response Theory to score responses, student-level rather than test-level reliabilities are generally reported. However, estimates of test-level reliabilities (Cronbach's alpha) for MAP Growth generally exceed .95.

Self-efficacy Survey. The self-efficacy survey consists of four Likert-scale items. Descriptive statistics for observed scores are presented for each year of the study in Table 3 and the specific items are provided in the Appendix. These observed scores were produced by taking the mean within-person score across all four items. Thus, with five possible response categories, plausible values for the within-person means ranged from one to five. Additional details on the survey can be found in West et al. (2017).

For modeling purposes, observed scores were not used. Rather, two approaches were employed, one in which models included a measurement submodel, another in which factor scores from models used to test for longitudinal measurement invariance were utilized. In general, results did not differ based on these two latent variable approaches, therefore only results using factor scores are reported. More details on the measurement model used are presented below.

Running Head: ELL Self-efficacy

Analytic Approach

For the first research question (Does self-efficacy differ for ELLs versus non-ELLs?), basic descriptive statistics were used. Specifically, mean self-efficacy and achievement test scores in reading and math were estimated by year for ELLs and non-ELLs. These mean scores were then plotted with time on the horizontal axis to show how much they differed between ELLs and non-ELLs at each timepoint.

To answer the second and third research questions on whether being an ELL is associated with growth in the three constructs (self-efficacy, math, and reading) and whether self-efficacy mediates the association between ELL status and growth, SEMs were used. Before using self-efficacy scores in a growth model, longitudinal measurement invariance tests were conducted to ensure the suitability of the scores for such uses. Details on those measurement invariance tests, including a path diagram for the model used, can be found in the Appendix.

After measurement invariance testing, the study employed conditional multivariate latent curve models (Bollen & Curran, 2006; Curran, Howard, Bainter, Lane, & McGinley, 2014). Figure 1(a) shows the multivariate SEM with joint growth models for math and self-efficacy. Per traditional SEM growth models, for the intercept random effects (α_{math} and α_{eff}), all loadings were set equal to one. With $t \dots T$ timepoints where $t = 1$ and $T = 4$ in the sample, the loadings on the growth random effects (β_{math} and β_{eff}) were constrained equal to $T-1$. In these models, the covariance matrix of the random effects was freely estimated. As described previously, these models were fit both with a measurement submodel and using factor scores. Given the results did not change between approaches and, thus, only models using factor scores are reported, the path diagrams do not include measurement submodels.

Running Head: ELL Self-efficacy

Figure 1(b) then shows how the model is adapted to include ELL status as a time-invariant covariate, as well as to consider different relationships among the latent variables germane to questions two and three. There are two primary changes to the path diagram in Figure 1(a). First, the four observed self-efficacy and math scores are now omitted from the path diagram for parsimony. Second, the covariances between the math intercept and self-efficacy slope, as well as the self-efficacy intercept and math slope, have been replaced with regression paths. Changing to regressions is not meant to suggest that the models are now causal in a quasi-experimental sense (Rubin, 2001). Replacing covariances with regressions simply represents a rescaling of the parameter and, as a result, the models are likelihood equivalent (Bollen & Curran, 2006; Curran et al., 2014). Thus, one cannot even test whether the model fits better with covariances versus regression paths. At the same time, this rescaling is useful when using time-invariant covariates like the ELL variable to determine whether the association between ELL status and the achievement latent growth parameter is mediated by the self-efficacy latent intercept (Bollen & Curran, 2006; Curran et al., 2014).

While estimates are not causal, the final models did control for a rich set of student- and school-level covariates likely associated with both achievement and ELL status. At the student level, the covariates included socioeconomic status (free or reduced price lunch designation), parents' highest degrees obtained, special education status, and biological sex (male/female). Although some of these controls like special education status are technically time-varying, student designations shifted minimally over the course of the sample. Therefore, the variables were treated as time-invariant covariates. At the school level, covariates included measures of poverty, federal funding for low-income students (Title I), and enrollment.

Running Head: ELL Self-efficacy

All models were initially fit separately by construct (math, reading, and self-efficacy). Within a construct, growth models were fit sequentially starting with a random intercept model followed by an intercept and slope model, then finally by a model with a quadratic term. Several fit indices were used to compare model fit across functional forms including the Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) (Bentler, 1990). Once the optimal functional forms were determined for each construct separately, model fit was examined for the multivariate models with achievement and self-efficacy included. Tables with fit statistics for models of self-efficacy scores only, achievement scores only, and both (as shown in Figure 1) are provided in the Appendix.

Results

Below, results are presented on a question-by-question basis. For Research Questions 2 and 3, results were generally insensitive to whether R-FEP students were include in the analyses in the ever-ELL category.¹ Therefore, only results that include them are reported.

Question 1. Does Self-efficacy Differ for ELLs Versus non-ELLs?

Figure 2 shows plots of mean self-efficacy, math, and reading scores by year and ELL status. Before turning to subgroup differences, one should note the steady decline in self-efficacy and incline in achievement between 5th and 8th grade. Thus, self-efficacy consistently declines for all students even while achievement grows steadily.

In general, there are gaps in math, reading, and self-efficacy favoring non-ELLs. T-tests suggest that all mean differences across constructs at a given timepoint are significant at the .01 level. To help understand the magnitude of the gaps in self-efficacy, scores were also standardized for this sample by year. As an effect size, the gaps in self-efficacy are quite large: in 2015, scores for ELLs versus non-ELLs differed by more than .35 standard deviations.

Running Head: ELL Self-efficacy

Returning to the figure, there also appear to be slight differences in trends for each construct over time by ELL status. The SEMs in the study allow those potential trend differences to be tested empirically.

Question 2. Do Growth Patterns in Self-efficacy and Achievement Differ for ELLs Versus non-ELLs?

Results suggest the model presented in Figure 1(b) fits the data well in both math and reading. In math, the RMSEA was .039 and the CFI was .995. In reading, the RMSEA was .045 and the CFI was .992. These statistics all suggest excellent model fit (Bentler, 1990). Details of these models and fit statistics can be found in the Appendix.

Figure 3 presents path diagrams with fully standardized coefficients in math and reading (using Mplus' StyYX command: for more details see Muthén & Muthén, 1998-2017). Before getting into comparisons of ELLs and non-ELLs, a couple of relationships bear mention. First, there is a significant and large association between the achievement and self-efficacy intercepts, as well as the achievement and self-efficacy slopes. At first glance, the high correlation between the slopes may seem counterintuitive given the finding in the first research question that achievement is going up just as self-efficacy scores are going down during the sample. To help illuminate this relationship, Figure 4 shows a scatterplot of empirical Bayes estimates of the two slope parameters. As the figure illustrates, although the slope coefficient on self-efficacy is often negative, there is still a positive association between the slopes. Finally, returning to the path diagrams in Figure 3, one should note that there is a significant association between the self-efficacy intercept and achievement slope.

Turning to the specific research question, according to the diagram, ELLs have math scores in 5th grade that are roughly .6 standard deviations lower than for non-ELLs. By contrast,

Running Head: ELL Self-efficacy

the coefficient on the path from ELL status to the math slope term is indistinguishable from zero, indicating that being an ELL is not associated with faster or slower growth in math. For self-efficacy, ELLs have initial scores that are .23 standard deviations lower, but ELL status is not significantly associated with the slope on self-efficacy. Thus, ELLs start out much lower on self-efficacy than their peers, and those scores do not appear to grow more over time relative to non-ELLs.

Turning to reading, ELLs have initial achievement scores that are .65 standard deviations below non-ELLs. However, being an ELL is associated with growth in reading between 5th and 8th grade that is .13 standard deviations higher than among their peers. On its face, this finding makes sense: as ELLs continue to master English, their academic proficiency in reading begins to catch up over time (Cook et al., 2011; Umansky & Reardon, 2014). The coefficients on ELL status for the self-efficacy intercept and slope are comparable to those for the math model. In the reading model, ELLs begin with self-efficacy scores that are roughly a quarter standard deviation below those of their peers, and the relationship between ELL status and the self-efficacy slope factor is negative but indistinguishable from zero. Thus, ELLs in the sample start with low self-efficacy and fail to make up ground despite faster growth in reading.

Question 3. Does Self-efficacy Mediate the Association between ELL Status and Achievement Growth?

Figure 3 also provides evidence on the relationship between growth in achievement and initial self-efficacy in 5th grade, including how that relationship differs by ELL status. In both math and reading models, the intercept in math does not appear to have any discernible association with the slope for self-efficacy. However, when regressing the achievement slope on the self-efficacy intercept, the coefficients are large and significant. In math, a one standard

Running Head: ELL Self-efficacy

deviation change in the student's self-efficacy intercept is associated, on average, with a .15 standard deviation increase in the student's growth trajectory in math. For reading, that coefficient is even higher at .19 units. Though not causal, this finding provides evidence consistent with a hypothesis that a student's initial self-efficacy is associated with the rate at which those students grow in math and reading over time.

Given these models are conditional on ELL status, the coefficients just described are for non-ELLs. As the figures show, there are differences in how the constructs relate for ELLs. In math, being ELL has no discernible direct effect on growth in math. However, ELL status has a strong indirect effect through its association with the self-efficacy intercept. The indirect effect can be calculated as $-.233 \times .153 = -.036$. In practical terms, given that being an ELL is negatively associated with self-efficacy in 5th grade, and self-efficacy at that time point is positively associated with growth in math, ELL status has a strong negative indirect association with growth in math estimated at .036 standard deviations. That indirect (mediating) effect is significant at the .01 level.

The story is similar in reading, though ELL status does have a positive direct effect on the reading slope estimated at .13 standard deviations. However, the indirect effect of ELL status on growth in reading via self-efficacy is $-.244 \times .187 = -.046$ standard deviations (also significant at the .01 level). According to these results, the gains that ELLs see in reading over time are tempered because they also tend to start with much lower self-efficacy.

Discussion

Closing achievement gaps for ELLs is a high priority in education policy and practice (Cook et al., 2012). ELLs generally have much lower math and reading achievement than their native-speaking peers, in part because they must learn core academic content in tandem with the

Running Head: ELL Self-efficacy

English language (Cook et al., 2011; Liguanti & Cook, 2013). As a result, ELLs are much more likely to drop out of high school than other students (Rumberger, 2006). Given ELLs start with lower achievement, which carries serious consequences for their long-term educational attainment, getting them caught up to their peers in math and reading is essential.

As shown in prior literature, academic self-efficacy (a prominent construct related to SEL) is a fundamental building block of motivation, and is therefore associated with achievement, including the growth needed to shrink achievement gaps (Bandura, 1993; Caprara et al., 2008). If students do not believe they can accomplish difficult academic tasks (low self-efficacy), there is little incentive to attempt them (Schunk & Meece, 2006). Given the academic challenges ELLs face are significant (Cook et al., 2011; Liguanti & Cook, 2013), one could imagine that self-efficacy is especially important for this subgroup.

This study produces evidence that the importance of self-efficacy to ELLs may be more than conjecture. First, there are large gaps in self-efficacy between ELLs and non-ELLs, as well as in math and reading achievement. In 5th grade, ELLs have self-efficacy scores that are roughly .25 standard deviations lower than for their peers. For math and reading the gap is even larger at roughly .6 standard deviations in both subjects. Thus, ELLs in the district studied begin middle school with not only an achievement gap, but also a self-efficacy gap.

Results further show that this gap in self-efficacy does not change much between 5th and 8th grade. In the SEMs used, the coefficient on ELL status is not significant when the dependent variable is the self-efficacy growth term in the model. That is, model results indicate that ELLs start with much lower self-efficacy, but do not appear to gain ground by the end of middle school. Notably, this self-efficacy gap largely remains stable despite there being a positive association between ELL status and growth in reading. Therefore, ELLs appear to gain no

Running Head: ELL Self-efficacy

additional confidence in their academic abilities despite having a slope on reading achievement that is .13 standard deviations higher than for non-ELLs.

Findings also indicate that being ELL has a significant indirect association with growth in math and reading through the gap in self-efficacy in 5th grade. According to the models, self-efficacy in 5th grade is associated with growth in math and reading, and thus the lower self-efficacy among ELLs in 5th grade is associated with growth in achievement for that subgroup (self-efficacy mediates the relationship between achievement and ELL status). In reading, being an ELL is associated with a .13 standard deviation increase in the slope on achievement. At the same time, that slope is also -.05 standard deviations lower because ELLs start with such low self-efficacy, which is associated with growth in reading. For math, there is no statistically discernible relationship between ELL status and growth, at least not directly. However, indirectly through 5th grade self-efficacy, that math slope is -.04 standard deviations lower for ELLs. This result, though not causal, is in-line with a theory that being an ELL reduces growth in math in part because that subgroup lacks self-confidence in the subject.

Beyond ELLs, these findings provide new evidence that may further justify the growing focus on SEL competencies in education policy and practice (Farrington et al., 2012). One reason for the increased emphasis on SEL is the association between related constructs and achievement, which could provide another option for addressing achievement gaps (Farrington et al., 2012; West et al., 2017). This study shows that, for at least one SEL-related construct and student subgroup, lower SEL scores are associated with decreased growth in math and reading. Therefore, a lack of belief among students in their academic abilities may be contributing to the slow rate at which achievement gaps are closed. In short, not only are SEL mindsets and competencies likely important to finishing high school and attending college (Farrington et al.,

Running Head: ELL Self-efficacy

2012); they may be a factor underlying the gaps in math and reading achievement that have been a focus of education practice and policy for decades. More research is needed to understand the full extent of the relationship between achievement gaps and SEL.

Limitations

There are limitations to this study that bear mention. The most notable relate to generalizability. The data come from a large urban district in California serving a high proportion of low-income ELL students. On one hand, this district is different from many others in the U.S. On the other, the district studied likely represents exactly the sort of context where questions of achievement gaps and self-efficacy for ELLs are of paramount importance. Further, the trends observed in self-efficacy, achievement, and their association have been documented elsewhere, including international contexts (Caprara et al., 2008; Farrington et al., 2012; West et al., 2017). Nonetheless, findings should be replicated in other settings and with other students.

Relatedly, due to slight differences in the survey instrument between elementary and middle school grades, this study focused on a cohort of students that began in 5th grade and finished in 8th grade. Therefore, results should be replicated with younger students to better understand the connections between ELL achievement gaps and self-efficacy in elementary grades. To check the sensitivity of results to the cohort used, the same estimates were produced for students starting in 6th grade. Results did not change substantively.

As previously noted, these analyses are also not causal. Rather, they use mediation analyses to test whether being an ELL appears to influence growth in achievement directly or indirectly via the association between growth and baseline self-efficacy. Thus, analyses are meant to provide insight into the mechanisms by which ELL status may affect academic growth.

Conclusion

Running Head: ELL Self-efficacy

ELLs face the daunting task of learning English and core academic content in math and reading at the same time. Partially as a result, ELLs tend to have lower mean achievement and high school graduation rates that are much lower (Rumberger, 2006). Given the long-term consequences, getting ELLs caught up on core content as they learn English—in essence closing achievement gaps between ELLs and non-ELLs—is important. This study provides evidence that these gaps might be closing more slowly because ELLs have less faith in their academic abilities than their peers. Further, self-efficacy for this group does not increase faster despite more rapid growth in reading among ELLs. That is, even when ELL achievement scores are growing faster than compared to native-speaking peers, ELLs' self-efficacy does not appear to be catching up. Though not causal, these findings are commensurate with the theory that finding ways to improve academic self-efficacy among ELLs could also lead to greater academic growth for those students, including closing achievement gaps with non-ELLs.

Notes

1. Though point estimates of the associations between ELL status and the intercept/growth terms for both achievement and self-efficacy did shift slightly depending on whether R-FEPs were included in the sample (Table 2, column 6), their significance and general magnitude remained unchanged, as did ultimate conclusions about the mediating effect of self-efficacy.

References

- Aguirre-Muñoz, Z., & Boscardin, C. K. (2008). Opportunity to learn and English learner achievement: Is increased content exposure beneficial? *Journal of Latinos and Education, 7*(3), 186–205.
- Bandura, A. (1994). Self-efficacy. *Encyclopedia of human behavior, 4*(4), 71-81.
- Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*(2), 117–148.
- Becker, B. E., & Luthar, S. S. (2002). Social-emotional factors affecting achievement outcomes among disadvantaged students: Closing the achievement gap. *Educational Psychologist, 37*(4), 197–214.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. 467). Hoboken, NJ: John Wiley & Sons.
- Britner, S. L., & Pajares, F. (2001). Self-efficacy beliefs, motivation, race, and gender in middle school science. *Journal of Women and Minorities in Science and Engineering, 7*(4).
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching, 43*(5), 485–499.
- Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal, 42*(2), 305–328.
- Callahan, R. M. (2013). The English learner dropout dilemma: Multiple risks and multiple resources. *Santa Barbara, CA: California Drop out Research Project.*

Running Head: ELL Self-efficacy

- Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology, 100*(3), 525.
- Castro-Olivo, S. M., Preciado, J. A., Sanford, A. K., & Perry, V. (2011). The academic and social-emotional needs of secondary Latino English learners: Implications for screening, identification, and instructional planning. *Exceptionality, 19*(3), 160–174.
- Cook, H. G., Linquanti, R., Chinen, M., & Jung, H. (2012). National evaluation of Title III implementation supplemental report: Exploring approaches to setting English language proficiency performance criteria and monitoring English learner progress. Washington, D.C.: *Office of Planning, Evaluation and Policy Development, US Department of Education*.
- Cook, H. G., Boals, T., & Lundberg, T. (2011). Academic achievement for English learners: What can we reasonably expect? *Phi Delta Kappan, 93*(3), 66–69.
- 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.
- Dabach, D. B. (2014). “I am not a shelter!”: Stigma and social boundaries in teachers’ accounts of students’ experience in separate “sheltered” English learner classrooms. *Journal of Education for Students Placed at Risk (JESPAR), 19*(2), 98–124.
- 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*

Running Head: ELL Self-efficacy

noncognitive factors in shaping school performance—A critical literature review.

Chicago, IL: Chicago Consortium on School Research at the University of Chicago.

Fast, L. A., Lewis, J. L., Bryant, M. J., Bocian, K. A., Cardullo, R. A., Rettig, M., & Hammond, K. A. (2010). Does math self-efficacy mediate the effect of the perceived classroom environment on standardized math test performance? *Journal of Educational Psychology*, *102*(3), 729.

Fry, R. (2008). *The role of schools in the English language learner achievement gap.*

Washington, DC: Pew Hispanic Center.

Gandara, P., & Orfield, G. (2010). A return to the "Mexican room": The segregation of Arizona's English learners.

Kanno, Y., & Cromley, J. G. (2013). English language learners' access to and attainment in postsecondary education. *Tesol Quarterly*, *47*(1), 89–121.

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.

Lewis, J. L., Ream, R. K., Bocian, K. M., Cardullo, R. A., Hammond, K. A., & Fast, L. A. (2012).

Con Carino: Teacher Caring, Math Self-Efficacy, and Math Achievement among Hispanic English Learners. *Teachers College Record*, *114*(7), 7.

Linquanti, R., & Cook, H. G. (2013). Toward a "common definition of English learner": A brief defining policy and technical issues and opportunities for state assessment consortia.

Washington, D.C.: *Council of Chief State School Officers.*

Running Head: ELL Self-efficacy

- Loeb, S., Soland, J., & Fox, L. (2014). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Educational Evaluation and Policy Analysis, 36*(4), 457–475.
- Matta, T. H., & Soland, J. (2018). Predicting Time to Reclassification for English Learners: A Joint Modeling Approach. *Journal of Educational and Behavioral Statistics*. Advanced online publication. <https://doi.org/10.3102/1076998618791259>
- Mercer, S. (2012). The dynamic nature of a tertiary learner's foreign language self-concepts. In *New perspectives on individual differences in language learning and teaching* (pp. 201-215). Heidelberg, Berlin: Springer.
- Niehaus, K., & Adelson, J. L. (2013). Self-concept and native language background: A study of measurement invariance and cross-group comparisons in third grade. *Journal of Educational Psychology, 105*(1), 226.
- Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24*(2), 124–139.
- Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86*(2), 193.
- Pompa, D., & Villegas, L. (2017). *Analyzing state ESSA plans for English learner accountability: A framework for community stakeholders*. Washington, D.C.: Migration Policy Institute.

Running Head: ELL Self-efficacy

- Randhawa, B. S., Beamer, J. E., & Lundberg, I. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. *Journal of Educational Psychology, 85*(1), 41.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2*(3–4), 169–188.
- Rubinstein-Avila, E. (2003). Facing reality: English language learners in middle school classes. *English Education, 35*(2), 122–136.
- Rumberger, R. W. (2006). *Tenth grade dropout rates by native language, race/ethnicity, and socioeconomic status*. Davis, CA: University of California Linguistic Minority Research Institute.
- Schunk, D. H., & Meece, J. L. (2006). Self-efficacy development in adolescence. *Self-Efficacy Beliefs of Adolescents, 5*, 71–96.
- Soland, J., Hamilton, L. S., & Stecher, B. M. (2013). *Measuring 21st-Century competencies: Guidance for educators*. Santa Monica, CA: The RAND Corporation.
- Soland, J. (2018). Can response process metadata provide insight into hard-to-measure constructs? Initial evidence from an international achievement test. Portland, OR: NWEA.
- Umansky, I. M., & Reardon, S. F. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal, 51*(5), 879–912.
- Usher, E. L., & Pajares, F. (2006). Sources of academic and self-regulatory efficacy beliefs of entering middle school students. *Contemporary Educational Psychology, 31*(2), 125–141.

Running Head: ELL Self-efficacy

Vuong, M., Brown-Welty, S., & Tracz, S. (2010). The effects of self-efficacy on academic success of first-generation college sophomore students. *Journal of College Student Development, 51*(1), 50–64.

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.*

Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29*(3), 663–676.

Running Head: ELL Self-efficacy

Table 1

Statistics on Analytic Sample

	2014-15	2015-16	2016-17	2017-18
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
Number of students with self-efficacy scores	2,319	3,266	2,842	2,694
Math percentile (median)	29	31	33	30
Reading percentile (median)	29	25	29	27

Running Head: ELL Self-efficacy

Table 2

Breakdown of Students Identified as "Ever ELL" in the Sample

Year	Non-ELL			Ever-ELL			Total
	EO/I-FEP (1)	LEP (2)	R-FEP (3)	EO/I-FEP (4)	LEP (5)	R-FEP (6)	
2014-15	1,270	0	0	0	1,062	0	2,332
2015-16	1,636	0	0	0	945	116	2,697
2016-17	1,381	0	0	0	644	219	2,244
2017-18	1,591	0	0	0	568	283	2,442

EO = English Only

I-FEP = Initial Fluent English Proficient

LEP = Limited English Proficient

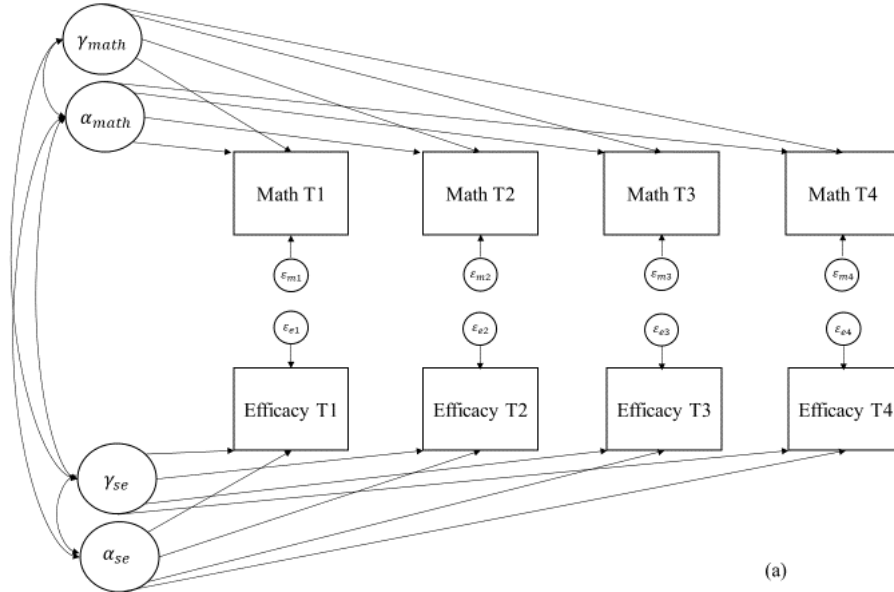
R-FEP = Reclassified English Proficient

Table 3

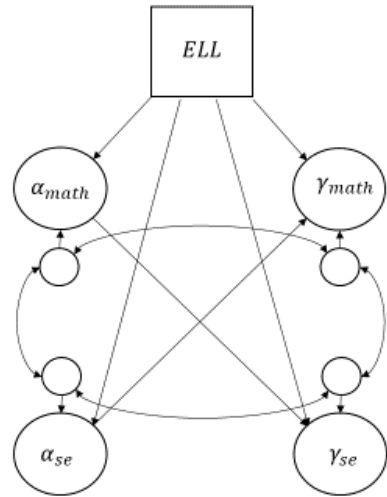
Statistics on SEL Survey Used by the District

	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

Running Head: ELL Self-efficacy



(a)



(b)

Figure 1. Path diagrams for multivariate latent curve SEMs.

Note. The “se” subscript refers to self-efficacy. “T1” refers to “Time 1” or the 2014-15 school year, just as “T4” refers to “Time 4” or the 2017-18 school year.

Running Head: ELL Self-efficacy

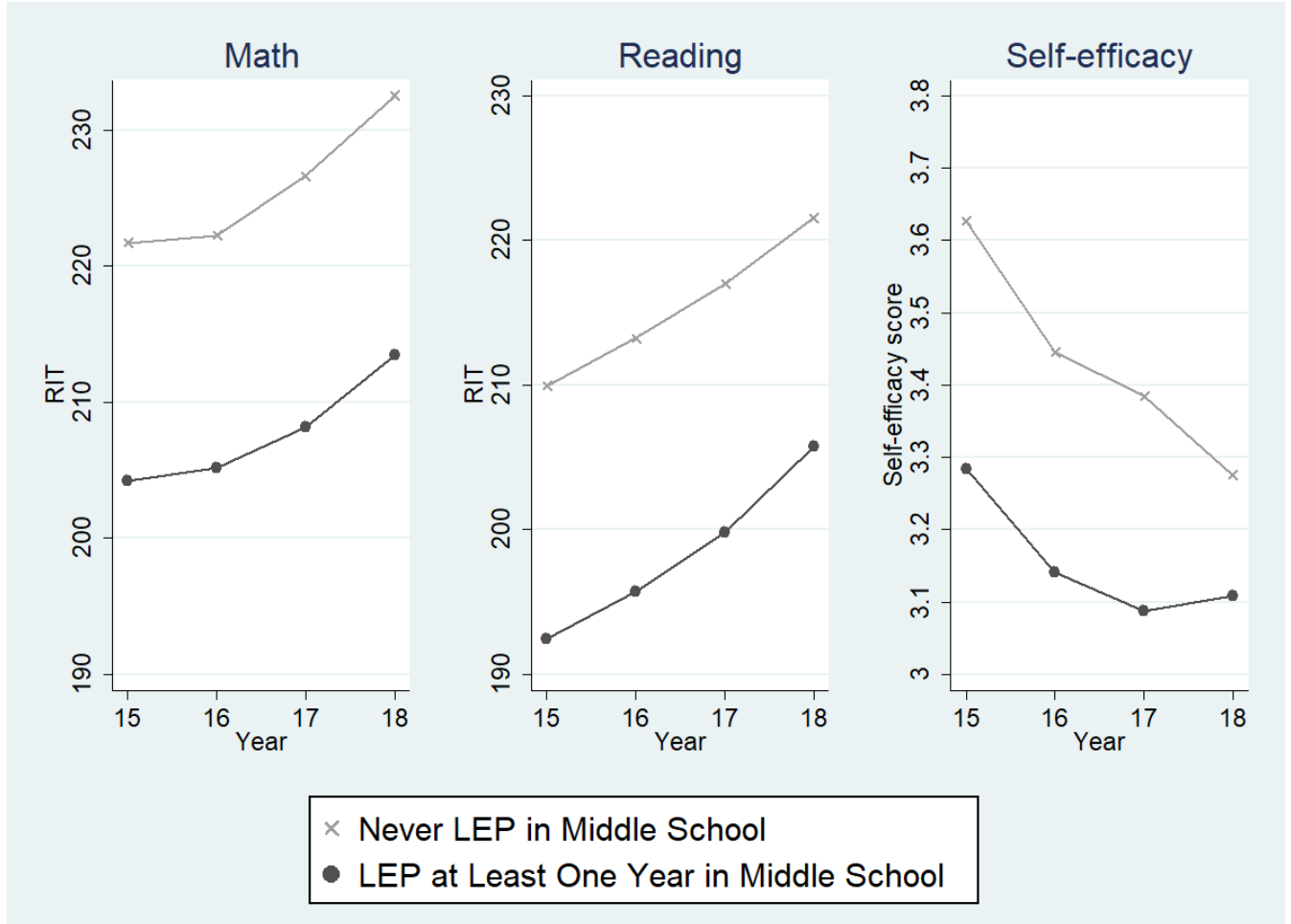


Figure 2. Scatterplot of math, reading, and self-efficacy scores by year (2014-15 through 2017-19).

Running Head: ELL Self-efficacy

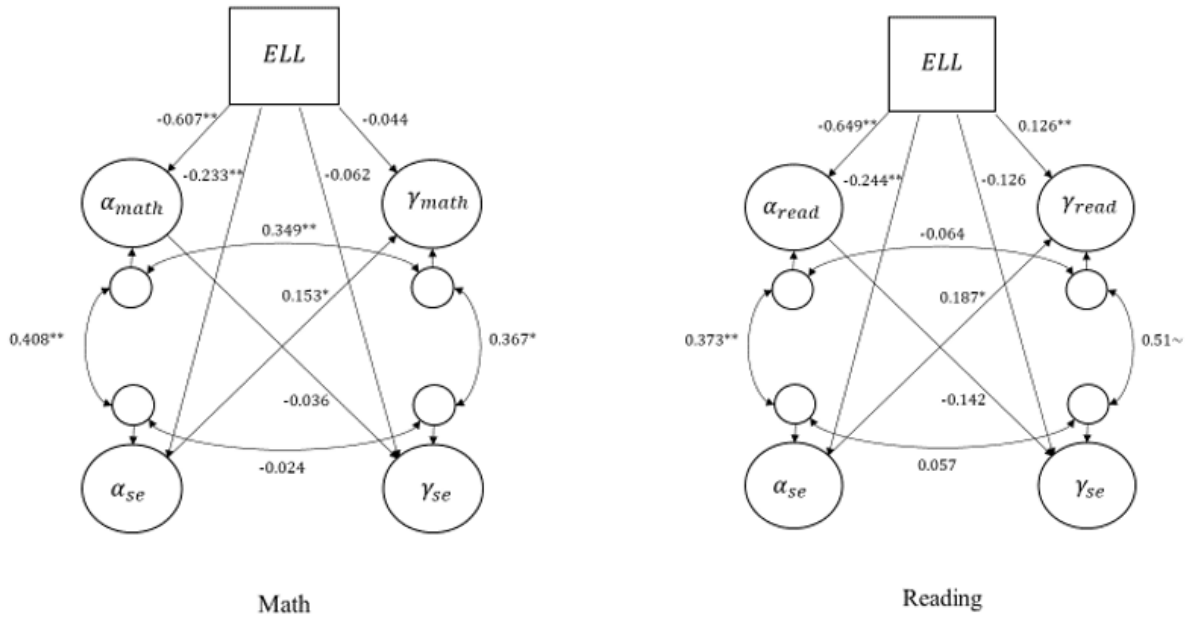


Figure 3. Path diagram from latent curve SEMs with fully standardized coefficients.

Note. Observed self-efficacy and achievement scores, as well as paths from observed scores to growth latent variables, omitted for parsimony. Statistical significance is denoted as follows: ** = .01, * = .05, ~ = .1.



Figure 4. Scatterplot of Multivariate SEM Slope Parameters in math and self-efficacy.

Appendix A. Survey Items

Table A1

Self-efficacy Items from District Survey

Agree or disagree with the following (5 point Likert scale)

I can earn an A in my classes.

I can do well on all my tests, even when they're difficult.

I can master the hardest topics in my classes.

I can meet all the learning goals my teachers set.

Appendix B. Model Fit Statistics

Table A2

Model Fit Statistics - Growth Models for Self-efficacy Scores

	Obs.	Test of Overall Fit			Likelihood Ratio Test			Fit Indices		
		χ^2	d.f.	p	$\Delta\chi^2$	d.f.	p	RMSEA	CFI	TLI
Model - Observed Scores										
M1. Intercept Only	4,157	328.948	8	0.000				0.098	0.844	0.883
M2. Intercept and Slope	4,157	43.118	5	0.000	285.830	3	0.000	0.043	0.981	0.978
M3. Intercept, Slope, and Quadratic	4,157	2.431	1	0.119	40.687	4	0.000	0.019	0.999	0.996
Model - Latent variable scores										
M4. Intercept Only	4,157	718.640	122	0.000				0.034	0.990	0.990
M5. Intercept and Slope	4,157	382.753	119	0.000	335.887	3	0.000	0.023	0.996	0.996
M3. Intercept, Slope, and Quadratic	4,157	316.772	115	0.000	65.981	4	0.000	0.021	0.997	0.997

Table A3

Model Fit Statistics - Growth Models for Math Achievement Scores

	Obs.	Test of Overall Fit		Likelihood Ratio Test		Fit Indices		
		χ^2	d.f.	$\Delta\chi^2$	d.f.	RMSEA	CFI	TLI
Model - Observed Scores								
M1. Intercept Only	4,037	4817.828	8			0.386	0.668	0.751
M2. Intercept and Slope	4,037	432.982	5	4384.846	3	0.146	0.970	0.965
M3. Intercept, Slope, and Quadratic	4,037	13.586	1	419.396	4	0.056	0.999	0.995

Note. All Chi-square and delta chi-square values significant at the .01 level.

Running Head: ELL Self-efficacy

Table A4

Fit Statistics for Multivariate Models of Self-efficacy and Achievement by Subject

Model	Obs.	Test of Overall Fit		Fit Indices				
		χ^2	d.f.	RMSEA	CFI	TLI	AIC	BIC
Math	4,157	104.691	14	.039	.995	.988	130678.619	130919.265
Reading	4,157	130.815	14	.045	0.992	.980	131190.785	131431.431

Appendix C. Measurement Invariance Testing

Before moving to bivariate LCM models, latent variable models designed to test for longitudinal measurement invariance of the self-efficacy scores were fit (Meredith, 1993). At each time point, latent variables were estimated based on each of the four self-efficacy item responses. Scale indeterminacy was addressed by constraining the first loading to equal one. Further, the variance of the first latent variable was constrained to be one, though variances of latent variables in subsequent time points were unconstrained. Each latent variable was allowed to correlate with the latent variables from other time periods. Correlations among observed item residuals for the same item in different time periods were also included (Curran et al., 2014; Widaman et al., 2010). Generally, this model follows the longitudinal measurement invariance approach described by Widaman, Ferrer, and Conger (2010).

This model was then used to test for measurement invariance over time. Oftentimes, measurement invariance is investigated across two groups of respondents. For longitudinal invariance, the groups become time periods to provide evidence the construct is being measured consistently over time (Widaman et al., 2010). The first test was for configural invariance, defined as having the same pattern of fixed and free factor loadings (and other parameters) across time, but no equality constraints across time (Koh & Zumbo, 2008; Meredith, 1993). Weak invariance was tested next by constraining the factor loadings for a given item to be the same across time periods. Third, strong invariance was examined by constraining the thresholds to be equal across time. Strict invariance was not tested because constraining the residuals to be equal across time is not always justifiable in a growth modeling context (Bollen & Curran, 2006). At each stage, a significant decrement in model fit was interpreted as a sign that

Running Head: ELL Self-efficacy

measurement invariance did not hold, and that the measures might not be suitable for longitudinal modeling (Kim & Yoon, 2011; Koh & Zumbo, 2008; Meredith, 1993).

Table A5 shows model fit statistics from measurement invariance tests. The configural model fits the data well, suggesting that there is likely a consistent factor structure across time. Further, imposing weak measurement invariance constraints actually improved model fit, though the LRTs are not significant at the .05 level. However, strong invariance constraints led to a statistically significant decrement in fit based on the LRT, though model fit is still excellent (RMSEA = .027). Upon further investigation, there were three threshold constraints that were responsible for the decline in fit between the weak and strong models. Therefore, removing those constraints (and thereby shifting to a partial measurement invariance model described by Meredith, 1993) helped improve the results. While the decrement in fit from the weak model to the partial strong model was significant at the .05 level, it was not significant at the .01 level. Given the model fit was excellent for partial invariance (RMSEA = .02), the partial model became the preferred model for subsequent analyses. Ultimately, these results indicate that fitting a growth model to the self-efficacy scores is justifiable.

Running Head: ELL Self-efficacy

Table A5

Model Fit Statistics from Longitudinal Measurement Invariance Tests of Self-efficacy

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	4,157	261.013	74	0.000				0.025	0.997	0.995
M2. Weak	4,157	250.775	83	0.000	13.416	9	0.145	0.022	0.997	0.996
M3. Strong	4,157	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)	4,157	296.922	114	0.000	220.340	12	0.000	0.020	0.997	0.997
Compared to Weak (M4 nested in M2)	4,157	296.922	114	0.000	49.583	31	0.018	0.020	0.997	0.997

Running Head: ELL Self-efficacy

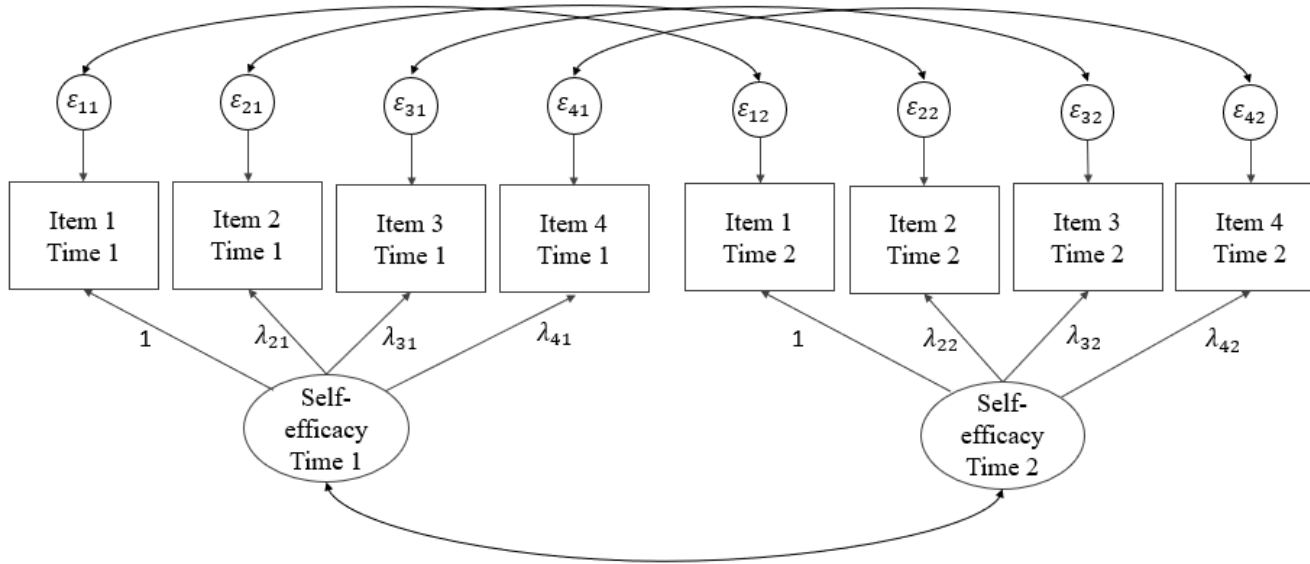


Figure A1. Path diagram for longitudinal measurement invariance testing model, two timepoints only.

ABOUT THE COLLABORATIVE FOR STUDENT GROWTH

The Collaborative for Student Growth at NWEA is devoted to transforming education research through advancements in assessment, growth measurement, and the availability of longitudinal data. The work of our researchers spans a range of educational measurement and policy issues including achievement gaps, assessment engagement, social-emotional learning, and innovations in how we measure student learning. Core to our mission is partnering with researchers from universities, think tanks, grant-funding agencies, and other stakeholders to expand the insights drawn from our student growth database—one of the most extensive in the world.

