EdWorkingPaper No. 22-636

Does Reclassification Change How English Learners Feel about School and Themselves? Evidence from a Regression Discontinuity Design

Monica Lee
Brown University

James Soland
University of Virginia

Reclassification can be an important juncture in the academic experience of English Learners (ELs). Literature has explored the potential for reclassification to influence academic outcomes like achievement, yet its impact on social-emotional learning (SEL) skills, which are as malleable and important to long-term success, remains unclear. Using a regression discontinuity design, we examine the causal effect of reclassification on SEL skills (self-efficacy, growth mindset, self-management, and social awareness) among 4th to 8th graders. In the districts studied, reclassification improved academic self-efficacy by 0.2 standard deviations for students near the threshold. Results are robust to alternative specifications and analyses. Given this evidence, we discuss ways districts might establish practices that instill more positive academic beliefs among ELs.

VERSION: August 2022

Suggested citation: Lee, Monica, and James Soland. (2022). Does Reclassification Change How English Learners Feel about School and Themselves? Evidence from a Regression Discontinuity Design. (EdWorkingPaper: 22-636). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/0t1s-a203

Does Reclassification Change How English Learners Feel about School and Themselves? Evidence from a Regression Discontinuity Design

Monica G. Lee

Brown University

Jim Soland

University of Virginia & NWEA

Abstract: Reclassification can be an important juncture in the academic experience of English Learners (ELs). Literature has explored the potential for reclassification to influence academic outcomes like achievement, yet its impact on social-emotional learning (SEL) skills, which are as malleable and important to long-term success, remains unclear. Using a regression discontinuity design, we examine the causal effect of reclassification on SEL skills (self-efficacy, growth mindset, self-management, and social awareness) among 4th to 8th graders. In the districts studied, reclassification improved academic self-efficacy by 0.2 standard deviations for students near the threshold. Results are robust to alternative specifications and analyses. Given this evidence, we discuss ways districts might establish practices that instill more positive academic beliefs among ELs.

Research increasingly demonstrates that student attitudes, perceptions, and emotions—often referred to as social-emotional learning or SEL skills—are salient factors in determining short- and long-term academic outcomes (Dweck et al., 2011). SEL skills consist of a wide range of attitudes, perceptions, and emotions that students need in order to meet academic goals, build positive social relationships, make good decisions, and express empathy for others (CASEL, 2005; Weissberg & Cascarino, 2013). As an example of the importance of SEL to long-term academic outcomes, students are unlikely to attend and persist in college if they cannot overcome setbacks and show the determination to meet long-term goals (Conley, 2008). For English Learners (ELs), resilience stemming from positive attitudes and perceptions is especially important because they face the daunting task of mastering the same academic content as their peers while also developing their skills in English concurrently (Goldenberg & Coleman, 2010). Yet, relatively little is known about SEL among ELs, especially how policies that affect ELs influence their social-emotional outcomes.

Given the academic challenges they face, ELs typically receive a range of educational supports to help them learn English while avoiding falling behind their peers in core content knowledge. A primary goal of federal education policy is to help ELs reach a point where they can succeed in school without those additional language supports, at which stage they are Reclassified as Fully English Proficient (RFEP). While such a goal is laudable, research shows that the effectiveness of such policies is highly contextualized. Much of the recent evidence on the effect of reclassification on academic outcomes, for example, has been mixed (Hill et al., 2014; Johnson, 2020; Pope, 2016; Robinson, 2011; Robinson-Cimpian et al., 2016). At the same time, there is evidence that, in certain contexts, reclassification can significantly impact achievement for reclassified students, both negatively and positively (Johnson, 2020; Shin, 2017;

Umansky, 2016b). In short, policies related to reclassification represent one potentially important policy lever to affect short- and long-term outcomes for ELs.

Though largely unstudied in quantitative settings, one could also imagine that reclassification (or failure to reclassify) impacts how ELs perceive themselves as learners. Several mechanisms could be driving potential effects of reclassification on SEL skills for ELs. For one, there is evidence that labeling students can lead to stigma (Rosenthal & Jacobson, 1968). For example, labeling effects have been demonstrated in the realm of federal education accountability by Papay, Murnane, and Willet (2011), who found that the achievement designations students receive under federal accountability (e.g., “below basic”) can impact student attitudes, such as college aspirations. Specific to ELs, there is extensive qualitative evidence that ELs feel stigmatized (Dabach, 2014) and devalued in the classroom despite their non-English linguistic abilities (Brooks, 2019; N. Flores et al., 2015). Some of these beliefs likely are compounded by the issue that English instruction targeted toward ELs often falls far below state standards, contributing to academic gaps that may persist during and after reclassification (Santibañez & Umansky, 2018).

Beyond labeling, there are other possible reasons that reclassification could impact ELs. For instance, one could imagine SEL being positively affected by reclassification if the student is academically prepared for mainstream instruction, but negatively affected if the student struggles academically when supports they might have still needed are prematurely removed (Robinson-Cimpian et al., 2016). Relatedly, the type of instruction students receive could affect SEL. For instance, some schools are implementing a co-teaching approach for ELD, with EL specialists "pushing in" to content-area classrooms (Honigsfeld & Dove, 2019). Such an approach could affect achievement for ELs, the perceived stigma associated with failing to reclassify, and how students perceive themselves as learners. Additionally, ELs are often placed in classrooms with low academic rigor and expectations (e.g., Callahan & Humphries, 2016; Dabach, 2014; Kanno & Kangas, 2014; Robinson-Cimpian et al., 2016; Shin, 2017; Umansky, 2016a), and reclassification may help to remove them from these settings and place them in learning environments that better support their academic self-beliefs. Despite these potential mechanisms for how reclassification could impact SEL skills, practically no quantitative research has examined related questions, let alone using quasi-experimental methods.

In this study, we examine the effect that reclassification has on ELs' SEL outcomes using data from several large districts. Specifically, we use a regression discontinuity (RD) design to examine how falling just above or below cut scores on tests used to determine reclassification affects students' SEL outcomes. Our study uses student surveys, administered to EL students across three school districts in California, that measure SEL skills across four constructs: growth mindset, academic self-efficacy, self-management, and social awareness. With this sample and methodological approach, we ask the following research questions: First, what is the impact of reclassification (i.e., removing EL status and support services) on a student's SEL outcomes? Second, does the impact of reclassification vary depending on a student's background or local context (e.g., biological sex or grade level)? As a preview of our findings, in the districts being studied, reclassification is associated (on average) with a significant increase in students' self-efficacy, suggesting that the EL label, level of academic content or academic expectations for ELs, or some mixture of these (and other) factors makes ELs feel less confident in their ability to succeed in school. While such findings are highly contextualized—that is, they may only apply to students in these districts during the years studied who are at the threshold for reclassification—they nonetheless provide plausibly causal evidence that reclassification can affect outcomes beyond achievement shown to be vital to long-term academic success (Almlund et al., 2011; Duckworth et al., 2007; Heckman & Kautz, 2013).

Background

In this section, we provide background on the educational context that ELs tend to experience. As part of that background, we describe what is known about reclassification and its impact on achievement. We then discuss the importance of SEL to academic outcomes, including for ELs.

EL Status and the Reclassification Process

The EL student population is growing across schools nationwide: Between 2000 and 2016, it increased by more than 1.1 million students (National Center for Education Statistics, 2019). Despite the prevalence of ELs, policies under the No Child Left Behind (NCLB) Act of 2001 designed to determine which students are considered ELs varied greatly depending on state, district, and local contexts. Since then, The Every Student Succeeds Act (ESSA) of 2015 has standardized many of these policies, including state-level guidance for entry and exit from EL services. Despite these changes between NCLB and ESSA, states and districts nonetheless maintain some latitude in how they determine which students are reclassified. Thus, even under ESSA, reclassification remains a viable policy mechanism for influencing EL outcomes.

Whether under NCLB or ESSA, students entering a school system (often in kindergarten) typically have their language proficiency assessed to determine EL status if they report using a language other than English at home, regardless of whether students are recently-arrived immigrants or were born in the United States (Abedi, 2008). These language proficiency tests usually measure reading, writing, speaking, and listening skills. Generally, students are designated as ELs if they do not score above a series of cut scores on tests of listening, speaking, reading, writing, and overall proficiency. Those who do pass the initial English test upon entering the school system are deemed to be Initially Fluent English Proficient (IFEP) and placed in mainstream academic settings with no additional supports.[1] Those who fall below one or more cut scores are usually deemed to be ELs.

Federal policy, in line with civil rights legislation, dictates that local agencies serve ELs by properly identifying them, offering language support services, and assessing this subgroup on both English proficiency and general academic content (Equal Educational Opportunities Act, 1974; Frey, 2018). Accordingly, ELs must receive programmatic services to succeed in both areas. First, ELs should receive English language development (ELD) instruction, the purpose of which is to improve their English language competency. In addition, ELs receive support in general academic content. In the state of California, where this study takes place, dual-language, transitional bilingual, and structured English immersion are three examples of programs offered to provide both of these supports to ELs (California Department of Education, 2021). Dual-language and transitional bilingual instruction both make use of English and a second language, but the degree to which the second language is used in the classroom varies greatly by program, as does the academic goal intended by each program. In contrast, structured English immersion serves ELs with English-only language support, as well as academic content catered toward students who are still gaining proficiency in English. For a full review on the different programs available to ELs, see Goldenberg and Coleman (2010).

Reclassification occurs when an EL is deemed to no longer need these special language and academic supports. Once a student is considered proficient in English across test-based

---

[1] This is a label specific to the state of California, and IFEP students may be simply referred to as “non-ELs” or a different label in other states.

criteria, they are eligible to become RFEP.[2] Upon being reclassified, ELs are placed into learning settings with non-EL peers (if they have not been already), and typically stop receiving comparable language support services. As previously described, language readiness is usually determined on the basis of whether students meet some criterion (cut score) on one or more English language proficiency (ELP) assessments. Several states also leverage a measure of academic readiness, such as the statewide primary achievement test in reading, in their reclassification decisions (Education Commission of the States, 2020). Further, school staff and parent/guardian consultations are additional qualitative criteria that 11 states currently use for reclassification (Villegas & Pompa, 2020).

Implications of EL Reclassification

Reclassification is typically associated with a change in instruction, as those who exit EL status are no longer entitled to receive additional English language support services. However, this varies drastically depending on the context within which ELs received language support services previously, and whether appropriate criteria were used for reclassification. For example, former ELs from pull-out programs may receive a greater proportion of mainstream classroom instruction than before upon reclassification, as ELs in these settings typically receive English language supports while their peers receive mainstream instruction (Goldenberg & Coleman, 2010). Other ELs, such as those in two-language environments, may experience fewer changes upon reclassification, as they continue to learn both core content and their native language in bilingual settings (Umansky & Reardon, 2014).

Additionally, reclassification can, at times, imply a change in the rigor of instructional content that ELs experience. In the state of California, where our own study is conducted, state

[2] Like IFEP, RFEP is a California-specific label, and students may be referred to as “former ELs” or a different label in other states.

law requires that reclassified ELs have full access to all “standardized instructional programs” that non-ELs also participate in, such as core subjects required for promotion/graduation (AB 2735). Given that EL coursework typically tends to be less rigorous than non-EL coursework (e.g., Callahan & Humphries, 2016; Dabach, 2014; Kanno & Kangas, 2014; Robinson-Cimpian et al., 2016; Shin, 2017; Umansky, 2016a), it is plausible that reclassified ELs may experience more rigorous instructional content than beforehand. This type of experience among ELs can hinge upon reclassification criteria, among other factors. If reclassification standards are set too low, students may struggle academically upon reclassification because the mainstream content is too challenging for them to master (Betts et al., 2019; I. Umansky et al., 2015). On the other hand, appropriate criteria and monitoring can help former ELs thrive academically upon transitioning to RFEP status (e.g., Robinson, 2011; Robinson-Cimpian & Thompson, 2016).

Perhaps understandably given the myriad ways reclassification can affect instruction and instructional settings, prior research demonstrates mixed evidence on the academic benefits of transitioning to RFEP status (Abedi, 2008; Callahan et al., 2010; Cimpian et al., 2017; Johnson, 2020; Robinson, 2011; Slama, 2014; Steele et al., 2017; Valentino & Reardon, 2017). Using an RD design, Robinson (2011) reported that the timing of reclassification could matter for achievement, although more recent studies have shown that the question of whether reclassification criteria are appropriately calibrated is highly specific to local contexts and policies (Betts et al., 2019; Pope, 2016). Evidence also suggests that reclassification, EL achievement, and the relationship between the two depend on a host of other factors, including the type of program/services offered to ELs (Umansky & Reardon, 2014), the grade level at which students are reclassified (Robinson-Cimpian & Thompson, 2016), the density of the EL student population in the locality (Bui, 2013; Callahan et al., 2009), each locality’s threshold for reclassifying (Cimpian et al., 2017), and implementation practices used to enact reclassification policies (Mavrogordato & White, 2017). In short, results examining the effects of reclassification on achievement are highly contextual, and one might not necessarily expect them to replicate across districts and studies.

The Importance of SEL Skills to Academic Outcomes

SEL skills include a range of attitudes, perceptions, and emotions that help "…children and adults understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions" (CASEL, 2005). While the makeup of SEL skills vary greatly depending on disciplines and settings, the range of skills within SEL includes intrapersonal (e.g., self-management), interpersonal (e.g., relationship-building), and cognitive (e.g., decision-making) skills – all essential for academic success and well-being (Dusenbury et al., 2015; Farrington et al., 2012; Soland et al., 2013). SEL skills are strongly linked to achievement and long-term outcomes (Almlund et al., 2011; Deming, 2017; Heckman et al., 2015; Heckman & Rubinstein, 2001). For example, students are less likely to attend and persist in college if they cannot overcome setbacks and show the determination to meet long-term goals (Duckworth et al., 2007).

Beyond academic outcomes, SEL skills also matter in their own right. Generally, developing strong SEL skills can reduce the probability of engagement in risky behavior or undesirable attitudes (Catalano et al., 2002). A recent meta-analysis found that the modal positive outcomes across 75 quasi-experimental or experimental studies included a decrease in antisocial behavior and an increase in social skills, showing the potential for schools to support SEL-related instruction that promotes healthy relationships and student well-being (Sklad et al., 2012). These skills are considered to be more malleable than traditional, cognitive skills such as performance on standardized assessments, which means school- and classroom-based strategies can influence SEL in meaningful ways (Almlund et al., 2011; Cunha & Heckman, 2006; Dweck et al., 2011; Farrington et al., 2012; Heckman & Kautz, 2013; Soland et al., 2019).

Our own study focuses on four constructs within the broader SEL umbrella: self-efficacy, self-management, growth mindset, and social awareness; Table 1 provides a brief definition and a sample survey item used to measure each construct. Self-efficacy is defined as a belief that students can achieve a given academic outcome; self-management is a self-perceived ability to regulate emotions, thoughts, and behaviors, especially in challenging circumstances; growth mindset refers to the belief that ability grows with effort; and social awareness is the ability to understand norms and empathize with others' perspectives (Bandura, 1977; CASEL, 2005; Dweck, 2006). In the Supplementary Online Materials (SOM), we define each of the SEL constructs considered in this study in greater detail. All four have been linked to positive academic and social outcomes (Duckworth et al., 2010; Duckworth & Seligman, 2005; Farrington et al., 2012; Zimmerman et al., 1992).

Differences in SEL Skills by Sociodemographic Factors

Extant literature presents substantial evidence on differences in student SEL skills by sociodemographic factors such as gender, economic disadvantage, and grade level of the student (Merolla, 2017; Schunk & Pajares, 2002; West, Pier, et al., 2018). Numerous studies examining SEL differences by gender in largescale contexts have demonstrated that female students report higher skills across all four SEL constructs in elementary school compared to male students (Claro & Loeb, 2019a; Fahle et al., 2019; Huang, 2013). However, there is evidence that, as students approach early adolescence, the gap in self-management, social awareness, and growth mindset between female and male students narrows (Claro & Loeb, 2019b, 2019a; Fahle et al., 2019; West, Pier, et al., 2018), indicating an interaction between age and gender. Further, Fahle et al. (2019) found that female students across five large, urban school districts reported significantly lower levels of academic self-efficacy and experienced steeper declines in this area of SEL in middle school relative to male students. Prior literature suggests that these trends may be attributable to differences in the biological development or in the socialization practices and gender stereotypes between boys and girls (Jacobs et al., 2002; Schunk & Pajares, 2002; Zimmerman & Martinez-Pons, 1990).

Turning to socio-economic status, most prior evidence suggests that economically disadvantaged students report lower average SEL skills compared to their counterparts. Students who live in economically disadvantaged neighborhoods report lower self-management skills (Papini et al., 1990), lower and more varying levels of academic self-efficacy (Merolla, 2017), and struggle with skills related to social awareness such as the ability to build relationships with their peers (Bolger et al., 1995). Recent large-scale studies examining student SEL have also found that factors that often proxy for socioeconomic status are significantly correlated with growth mindset, in that students who are eligible for free lunch or have parents with less than a college education were more likely to report fixed mindsets relative to their peers (Claro & Loeb, 2019b; Destin et al., 2019; Snipes & Tran, 2017; West, Pier, et al., 2018).

Lastly, studies provide mixed evidence on certain SEL skills by age groups. Generally, studies consistently suggest that self-management (e.g., Duckworth et al., 2010; West, Pier, et al., 2018) and self-efficacy (e.g., Pajares et al., 2000; Schunk & Meece, 2006; Schunk & Pajares, 2002) both decline in middle school during the onset of early adolescence (Soland et al., 2022). The findings on growth mindset are mixed (Pintrich & Zusho, 2002; West, Pier, et al., 2018;

Wigfield et al., 1996) and the array of evidence on this construct by age is limited, as much of the growth mindset literature focuses on the relationship between growth mindset and academic achievement rather than the construct in and of itself. Lastly, while social awareness—and skills associated with social awareness competencies, like self-awareness—can increase with age (Choudhury et al., 2006; Wigfield et al., 1991, 1996), other studies have shown the opposite (Gaspar, Tania, Cerqueira et al., 2018; West, Pier, et al., 2018). Regardless, there is evidence that SEL skills can differ by age, gender, and on the basis of socioeconomic factors.

SEL Skills Among ELs

As previously discussed, reclassification can affect a student's academic outcomes and learning setting, both of which are related to SEL outcomes like self-efficacy (Bandura, 1977; Merolla, 2017). However, there are additional factors related to being an EL that could impact SEL skills, including during the transition to RFEP. Extant literature exploring the 'subtractive schooling' environment (Valenzuela, 1999) into which ELs are initially placed suggest that several factors may contribute to differential attitudes and beliefs among ELs. First, there is evidence the EL label can be stigmatizing. Although ELs possess assets such as bilingualism and cultural funds of knowledge (González et al., 2006), educational systems define and distinctively label them by their lack of English proficiency, which can be stigmatizing to students (Dabach, 2014). This stigma may be worse for "long-term ELs", or students who have been designated as EL for more than five years (Estrada & Wang, 2018), as high-performing students move on from their EL status to mainstream settings and low-performing ones remain in ELD coursework (Abedi, 2008; Gándara et al., 2005). There is further evidence that ELs placed in ELD classes in secondary schools can feel even more stigmatized because those classes can appear (and often are) more remedial during high school (Thompson, 2015).

Second, ELs' academic self-beliefs may be influenced by the inequitable learning environments they can be placed in (Callahan, 2013; Parrish et al., 2006; Reardon & Galindo, 2009; Rumberger & Gándara, 2004). Although the purpose of EL status designation is to provide additional supports and services for students, required ELD coursework often supplants enrollment in other core subjects (Valdés, 1998). Thus, ELs can be excluded from more rigorous coursework (Callahan, 2013; Callahan et al., 2010; Callahan & Humphries, 2016; Dabach, 2014; Kanno & Kangas, 2014; Robinson-Cimpian et al., 2016; Shin, 2017; Umansky, 2016a). In the long run, ELs can have limited access to rigorous curricula or courses that will prepare them for college, and can subsequently lag behind in college preparatory course-taking (Callahan et al., 2010) as well as college attainment. Those who do attend college are more prone to enroll in remedial coursework, attend a two-year institution rather than a four-year institution, or both (Callahan & Humphries, 2016; S. M. Flores & Drake, 2014).

Third, expectations are often lower for ELs. Educators in EL-prevalent classrooms can, in some instances, expect less of EL students academically (Dabach, 2014), and there is evidence students can learn and absorb those lower expectations themselves (de Boer et al., 2018). In the pursuit of English proficiency, educators may position themselves as purely responsible for an EL's mastery of the English language more so than a comprehensive set of behaviors and beliefs that lead to academic success (Yoon, 2008). Indeed, studies suggest that ELs perceive that their teachers and families have lower expectations of them than of their non-EL peers (Dabach, 2014; Umansky & Dumont, 2021; Valdés, 1998). Generally, educators are more likely to have higher goals for and expectations of their high-track classrooms compared to low-track ones (Oakes, 1985; Raudenbush et al., 1993), and so educators may possess lower academic expectations of

EL students, especially if they come from a disadvantaged or minoritized race/ethnic student backgrounds or are shut out of rigorous coursework (Valenzuela, 1999).

One could imagine ELs factoring in their stigmatizing experiences, exclusion from advanced courses, and lower expectations from those around them into their own perceptions, beliefs, and attitudes about themselves as learners. Indeed, prior research suggests ELs have more negative academic attitudes and self-beliefs compared to non-ELs (Dabach, 2014; Valdés, 1998). Recent studies have found that ELs are more likely to have fixed mindsets about their academic success (i.e., the belief that one's intelligence cannot change with effort) relative to their peers (Claro & Loeb, 2019b; West, Pier, et al., 2018). LeClair et al. (2009) compared survey data among 257 elementary school students and found that EL students rated themselves lower in academic self-efficacy, and perceived their non-EL peers as better behaved and more able to follow classroom rules than themselves.

Some researchers also suggest that SEL skills can mediate the association between EL status and achievement. Niehaus and Adelson (2014) used a structural equation modeling (SEM) strategy to show that ELs' self-beliefs in math were positively and significantly correlated with their math achievement, and that Hispanic ELs had lower average self-beliefs than non-ELs. Meanwhile, Soland and Sandilos (2020) used a similar modeling strategy, showing that low self-efficacy is associated with slower academic growth in math and reading among ELs. Soland and Sandilos (2020) further conjectured that low academic self-confidence may be contributing to the achievement gap between ELs and non-ELs.

Conclusion

In conclusion, there is widespread evidence that ELs have lower scores on SEL measures due to a variety of factors, including lower expectations and stigma. Further, lower self-efficacy among ELs is associated with lower achievement and slower academic growth, possibly contributing to achievement gaps and the rate at which they close over time. Factors associated with reclassification like shifts in academic content/delivery could further affect SEL skills as student transition to RFEP status. Yet, little is known about the effect of reclassification on SEL outcomes. Understanding the causal effect of EL status on such outcomes in the particular context we study could provide educators and policymakers with needed information on the potential consequences of reclassification, including how EL status makes those students feel about their own academic capabilities. In the remainder of the study, we use RD to investigate the relationship between reclassification and SEL skills.

Methods

Sample

This study leverages data from the California CORE districts, a statewide network of districts that formed in 2010 to build an innovative data-sharing and accountability system. Beginning in 2014, the CORE districts implemented an annual survey to measure four SEL skills among its students: growth mindset, self-efficacy, self-management, and social awareness. We focus on three of the CORE districts, all of which fall within the top ten largest school districts statewide. The proportion of ELs in all three districts exceeded the state average (i.e., ELs constituted more than 21.6% of total enrollment) in the year studied. Student-level EL data from the CORE districts are merged to student demographics, academic achievement, and SEL survey responses to create our final analytic sample.

Figure 1 presents a simplified timeline for individuals in the study. The analysis follows a sample of students in grades 2 through 7 who were formally classified as ELs in the 2012-2013 school year. Then, we track reclassification in the following year (i.e., 2013-2014) and gauge the subsequent impact of reclassification via student SEL surveys that were completed in 2014-2015. We used this time period because ESSA led to considerable shifts in the tests used to reclassify students and associated policies as states transitioned away from NCLB. Thus, using a slightly earlier time period helped ensure we were studying effects of a consistent policy.

In the state and timeframe pertaining to our analysis, reclassification decisions were based on test-based criteria and two qualitative assessments.[3] The test-based criteria consisted of cut scores from two assessment batteries. The first included four subtests measuring speaking, listening, reading, and writing proficiency, as well as a composite of the four assessing overall proficiency, that comprise the California English Language Development Test (CELDT). The second included the English Language Arts (ELA) scores from the California Standards Test (CST). Qualitative assessments included input from teachers and parent/guardian consultation. Our study focuses on the test-based assessments that determine a student's probability of being reclassified. Each CELDT test is scored on one of five levels—Beginning (1), Early Intermediate (2), Intermediate (3), Early Advanced (4), and Advanced (5)—and the minimum threshold for reclassification is Early Advanced for the CELDT overall portion and Intermediate for the four subdomains of the CELDT. Scaled scores to reach these levels vary by grade level and subsection. For the ELA portion of the CST, while the California Department of Education required a minimum threshold of Beginning (300 points), the districts in this study established a minimum score that was higher than the state-mandated threshold as a means to ensure reclassified ELs were able to participate successfully in mainstream classrooms (Hill et al., 2014). For the purposes of analysis, we standardize each student's test score by centering on its

[3] The test-based criteria have since changed to reflect implementation of two new standardized assessments, the English Language Proficiency Assessments for California (ELPAC) and Smarter Balanced.

respective cut score and dividing by the sample-specific grade-level standard deviation (Reardon & Robinson, 2012).

Table 2 presents descriptive statistics on student demographics, achievement, and SEL skills for all ELs, as well as ELs separately by district. Because a regression discontinuity analysis examines students at the threshold, it is not reasonable to present descriptive statistics of all students in the district, as these statistics may not necessarily reflect those of students closer to the threshold. Therefore, the statistics presented are of ELs whose binding scores are within 1.5 standard deviations of the cut point, as determined by the CELDT and CST-ELA. There are patterns evident across all three districts in the sample. For example, students in the analytic sample are overwhelmingly Latinx with a small group of Asian students. Across the analytic sample, the majority of students who reclassify do so in the elementary grades, with the reclassification rate declining for students in older grades. Table 2 also presents average pre-treatment achievement scores, which are used to determine reclassification status later on in the analysis. The average CST-ELA score is below the cut point, which is to be expected given that the majority of the analytic sample consists of ELs who did *not* reclassify during the span of our study.

There is also descriptive evidence of heterogeneity by district. For instance, District 2 has fewer students from socioeconomically disadvantaged backgrounds relative to Districts 1 and 3, and has a lower reclassification rate. Additionally, District 3 has much higher average CELDT scores relative to Districts 1 and 2 and, accordingly, has a higher reclassification rate in the year of this study. The final analytic sample consists of students with nonmissing values on CELDT, CST-ELA, and SEL survey scores, as well as test scores falling within the optimal bandwidth, which is calculated separately for each SEL outcome.

Measures

The SEL survey administered to students in CORE districts measures four social-emotional constructs annually. The survey measures self-management (9 items), growth mindset (8 items), self-efficacy (4 items), and social awareness (8 items); a complete questionnaire can be found in the SOM. Students responded to items by choosing from a five-point Likert scale. The same survey was implemented across all students and districts in a given school year.

Rather than use scores from individual survey items or mean item scores across the four constructs as outcome variables, we use item response theory (IRT) scale scores to produce latent variable estimates of the desired constructs that account for measurement error. This approach was taken given research showing that sum/mean scores make large, oftentimes untenable assumptions (McNeish & Wolf, 2020) that can severely bias treatment effect estimates (Authors, 2021). Specifically, we use scores produced using a generalized partial credit model (GPCM). This approach has several advantages, including accounting for the “difficulty” of each item as well as the item’s association with the construct being measured (Muraki, 1992; Samejima, 1969). Extensive analyses of psychometric properties of the SEL survey items have found high reliability for item responses associated with each construct and provided evidence to support the use of the GPCM scaled scores for research (Meyer et al., 2018). Subsequently, the present analysis measures the impact of reclassification on GPCM scaled scores of the SEL constructs, which have a mean of 0 and a standard deviation of 1.

Binding Score Regression

One complication with establishing a regression discontinuity for EL reclassification is that the determination is made on multiple tests and associated cut scores, rather than a single assessment. To address this issue, we use the “binding score” regression discontinuity approach adopted by Robinson (2011) and enumerated by Reardon and Robinson (2012). This approach establishes a single rating variable from the multiple assessments considered for reclassification. We define a rating variable, r_i, to equal the minimum value of six standardized assessment scores for reclassification in the year we study for a student i, centered around their respective cut points:

$$r_i = \min\left(ELA_i, CELDT\ overall_i, CELDT\ listening_i, CELDT\ speaking_i,\right.$$
$$\left. CELDT\ reading_i, CELDT\ writing_i\right)$$

Thus, our RD analysis estimates the effect of reclassification for those whose lowest score among the five pre-reclassification tests was above zero and the student would therefore be expected to reclassify. We leverage this approach to estimate a given student's probability of receiving the treatment.

Analytic Approach

As mentioned previously, not all ELs scoring above the test cut scores are reclassified; other criteria such as teacher and parent/guardian input are also factors considered for reclassification, which implies that it is possible for some students, despite meeting all the test-based criteria, to remain ELs and fail to reclassify. The reverse is also possible, where some students reclassify despite not having met all the test-based criteria. Thus, we use a fuzzy RD design with the binding score as an instrument for being reclassified in order to address imperfect compliance. We employ the following two-stage least squares (2SLS) equations to estimate a treatment effect where, for student i in school district d:

$$reclass_{id} = \beta_0 + \beta_1 r_{id} + \beta_2 z_{id} + \beta_3 r_{id} * z_{id} + \gamma_d + e_{id}$$

$$y_{id} = \delta_0 + \delta_1 \widehat{reclass}_{id} + \delta_2 r_{id} + \delta_3 r_{id} * z_{id} + \gamma_d + u_{id}$$

Here, y_{id} is the outcome (each of the four SEL constructs estimated separately) for individual i; $reclass_{id}$ is a dummy variable which equals 1 if the individual receives treatment (i.e., was reclassified and no longer deemed EL) and 0 otherwise; z_{id} is a dummy variable which equals 1 if the individual has a binding score that is at or above the cut point and 0 otherwise; r_{id} is our rating variable, which is a continuous binding score generated from the multiple assessments used for reclassification and shows an individual's nearest distance from the threshold; and e_{id} and u_{id} are random error terms at each stage. Notably, whether or not a student passed all assessment cut points is used to predict whether they are reclassified, and their predicted probability of reclassification (ranging from 0 to 1) is used in the model to estimate the impact on students' self-perceived SEL skills. Lastly, the analysis includes district fixed effects, γ_d, across both the first and second stage equations, as reclassification policies are district-specific and results could vary meaningfully across districts.

From a modeling perspective, RDs are estimated using local linear regressions of SEL on the test score in question for students just above and below the cutoff. As recommended by Lee and Lemieux (2010), we consider flexible forms of the regression, allowing for different slopes on either side of the cut and using both linear and quadratic terms of r_{id}. After comparing model fit, the preferred model for this analysis is one with a linear spline on both sides of the threshold rather than higher-order polynomials, which can overfit the data (Gelman & Imbens, 2019). In addition, it includes district fixed effects γ_d, both to account for time-invariant district differences and because their inclusion considerably improves upon the strength of the first stage. All subsequent analyses, including heterogeneity analysis and robustness checks, also make use of linear rating variables and include district fixed effects.

A critical question when estimating an RD is how far an observation must lie from the cut score in order to be included in the estimation. With a sample of students who are too close to the cut score, the estimates run the risk of being imprecise and underpowered. With a sample that is too far from the cut score, our parameter estimates may not be accurately reflecting the impact of reclassification at the threshold. To address this bias-precision tradeoff, we use the cross-validation procedure developed by Imbens and Lemieux (2008), which optimizes the distance (or "bandwidth") relative to that tradeoff. We calculate Imbens and Kalyanaraman (2012) optimal bandwidths separately for each of the four outcome variables. Another popular strategy used to calculate optimal bandwidths with robust confidence intervals, developed by Calonico, Cattaneo, Farrell, & Titiunik (2017), yields similar results, as seen in the SOM (Table S1).

Robustness Checks

Through an array of robustness checks, we provide evidence that the RD design is valid and that our results remain robust to various specifications. First, we check for manipulation around the cut score by visualizing the distribution of test scores using McCrary's density test. Manipulation around the cut scores is unlikely but not impossible; the thresholds to qualify for reclassification for the CST-ELA and CELDT subdomains are known, do not change every year, and are available publicly. Nonetheless, the multiple-choice portions of the standardized assessments are first graded as raw scores (i.e., the proportion of correct answers out of all questions) and then converted to IRT scale scores, while the open-response portion of the CELDT was graded by external contractors (*California English Language Development Test: Technical Report*, 2015). Therefore, it would not be easy to gauge how many questions a student needs to score correctly in the *raw* form in order to push them over the *scale score* threshold. Nonetheless, we use a McCrary's density test to examine the issue.

Beyond potential manipulation, we conduct other sensitivity analyses common in the RD literature. The second sensitivity analysis we conduct is to use multiple bandwidths (20 alternative bandwidths) and examine whether results are robust. Third, we use a fuzzy frontier regression discontinuity recommended by Reardon and Robinson (2012) to account for the multiple tests used in the reclassification determination. Finally, we look for spurious discontinuities for variables unrelated to SEL at the cutoff (e.g., demographic variables) and for discontinuities in SEL scores away from the cut score.

Examining How High Correlations among SEL Outcomes might Affect Results

Beyond how the RD is specified, another potential concern is that we find effects of reclassification on a given SEL outcome not because there is a causal relationship, but because the specific outcome in question is associated with another SEL outcome that is impacted causally. For example, one might worry that reclassification appears to positively impact growth mindset when, in fact, the causal effect is on self-efficacy, which is highly correlated with growth mindset. This issue is especially important given SEL outcomes have often been shown to be highly correlated, including for the surveys we use (Soland et al., 2019; West, Pier, et al., 2018). To examine whether such a phenomenon might be occurring in our data, we began by regressing a given SEL score on the other three SEL scores (e.g., growth mindset regressed on self-efficacy, self-management, and social awareness). We then re-ran the RDs, but using the residuals from those regressions such that we could tell if reclassification had an effect after accounting for associations among the SEL constructs. Note that such an approach would likely lead to an understatement of true reclassification effects if becoming RFEP actually improves multiple SEL outcomes causally.

Results

We visually inspect the first stage in Figure 2 by plotting the relationship between average rates of reclassification, our treatment variable, and binned binding scores, binned at intervals of 0.05 SD. We observe a distinct jump at the cut point: students whose binding scores are in the bin immediately to the right of the cut point are approximately 40 percentage points more likely to be reclassified than those whose binding scores are in the bin that falls immediately to the left of the cut point. For ease of interpretation, we also visualize the first stage separately for the two assessments that comprise the binding score, and each respective jump in reclassification rates appears similar for these subsamples as the one observed in Figure 2 (see SOM, Figure S1). Our F-statistics confirm the visual evidence we have from the plots: all F-statistic values far exceed 10, with those from the linear model above 1440 and those from the quadratic model above 475 (see statistics in Table 3).

The coefficients in Table 3 are our main results of reclassification effects on SEL outcomes, with Column 2 showing results from our preferred model containing district fixed effects. At the optimal bandwidth, a model assuming a linear functional form and district fixed effects yields positive effects of reclassification on self-efficacy and growth mindset (see Column 2). Notably, at the cut point, reclassification appears to improve self-efficacy by 0.208 SD and growth mindset by 0.146 SD. The estimates for self-management and social awareness are positive as well: at the threshold, self-management and social awareness increase by 0.131 SD and 0.119 SD, respectively, upon reclassification. However, the impact of reclassification on self-management is positive only at the 10% significance level, and the estimate on social awareness is not statistically significant. A visual depiction of the discontinuity estimates in Table 3 can be seen in Figure S2.

Estimates of the impact of reclassification on the four SEL measures remain similar even when examining the impact of reclassification by subgroup. Our analytic strategy in exploring heterogeneity of results is identical to what is used in our preferred model in Table 3, except that we narrow the sample further to smaller subsamples of students by gender (male or female), economic disadvantage (eligible/ineligible for free or reduced-price lunch), and timing of reclassification (in elementary or middle school). Specifically, we observe gains in self-efficacy and growth mindset manifesting across most subgroups and in self-management and social awareness for a few subgroups. Generally, increases in self-efficacy appear similar upon reclassification for students regardless of eligibility for free or reduced-price school lunch (Table 4, Columns 3 and 4). However, female students and students who are reclassified in the elementary school grades appear to manifest greater increases in self-efficacy relative to their peers, a notable finding given prior descriptive evidence that female students in the specific elementary school grades studied (4-5th grades) typically experience a decline in academic self-efficacy (Fahle et al., 2019). Meanwhile, increases in growth mindset appear higher among male students, students who are not eligible for free or reduced-price school lunch, and students in elementary grades.

Results on Robustness Checks

Although the RD design requires relatively few assumptions to make a causal claim, there are nonetheless circumstances that can lead to erroneous conclusions. The following robustness checks show few notable issues with the validity of our RD design. First, the RD design assumes that there must be no manipulation of the rating variable at the threshold. Therefore, we use a McCrary's (2008) density test to examine this issue, and find no evidence of discontinuities in the density at the cut score for both the CST and CELDT overall tests (Figure S3), suggesting no manipulation.

Second, we examine sensitivity to bandwidth. This is particularly salient for our analysis, as our main results in Table 3 yield estimates above zero across all four SEL constructs. Thus, we estimate the impact of reclassification on self-efficacy using 20 additional bandwidths that are larger and smaller (in 0.02 SD increments) than the optimal bandwidth. Coefficients and confidence intervals for self-efficacy derived from these 20 additional samples, in addition to the sample used in the main model containing our optimal bandwidth, are shown in Figure 3. Across all 21 different samples, coefficients for self-efficacy are positive and statistically significant, showing the stability of the positive estimate on self-efficacy despite distance from the threshold.

At the same time, we find that reclassification effects on the other three SEL outcomes are sensitive to alternative bandwidth specifications (see Figure 3). For instance, estimates in our main analysis within the optimal bandwidth indicated statistically significant increases in growth mindset and self-management and a positive, nonsignificant increase in social awareness as a result of reclassification (Table 3). However, re-estimating the model at different bandwidths shows that the positive estimates on both growth mindset and self-management are no longer significant when using much larger bandwidths. This result suggests that, while there *may* be increases in students' perceived growth mindset and self-management after exiting EL status, the results could be less generalizable to a broader population of students beyond those closest to the threshold. Additionally, for all but one of the 20 bandwidths, the effect on social awareness was not significant at the .05 level.

Third, as an additional robustness check recommended by Reardon and Robinson (2012), we use a "fuzzy frontier regression discontinuity" approach. This approach compares students at the cutoff using only a single assessment as the instrument for treatment assignment, rather than a binding score, which takes into account all assessments/subtests. This approach also produces similar point estimates for self-efficacy (see Table S2).

Fourth, we look for spurious discontinuities. We begin by checking for jumps in pre-treatment covariates near the cut point, indicating that certain demographics may be sorting into the treatment, and that increases in SEL might be spurious. Figure S4 presents a visualization of pre-treatment, time-invariant demographic characteristics such as gender, race/ethnicity, and socioeconomic status across the binding score distribution. These covariates appear smooth at the threshold and are unrelated to the binding score, providing evidence that there is no systematic sorting of students of certain demographic characteristics to either side of the threshold. As a parallel analysis, we also use these demographic characteristics as the dependent variable in our 2SLS models, and find no statistically significant effect of reclassification on the likelihood of being female, Latinx, or economically disadvantaged (see Table S3 for estimates).

However, results from a placebo RD design, which moves the cut point to an arbitrary point away from zero, did produce some unexpected results. Specifically, we estimate our main model using two randomly chosen placebo cut points, set at one SD above and below the actual cut point (see Table S4)**.** When examining a placebo location below our actual cut point, while most estimates are small and indistinguishable from zero for all SEL outcomes, growth mindset shows a significant jump for students close to the new, arbitrary threshold. When conducting the same placebo test one SD above our true cut point, we find nonzero estimates for self-efficacy, growth mindset, and social awareness, although none of these are statistically distinguishable from zero at the .05 level.

Upon first glance, these estimates may be cause for concern, as in theory a placebo test should indicate smoothness across the entirety of the rating variable, save the cut point. However, visualizations of average SEL scores (as seen in Figure S2) already indicate that estimates of means at the higher end of our binding score vary considerably more by bin, which may be affecting the results we observe in our placebo test. That is, scores at the extremes of the binding score are highly variable, especially compared to near the cut score, such that arbitrarily setting a new threshold at an extreme could easily make jumps occurring due to random variability look meaningful. Estimates in Column (2) are also from models with far smaller F-statistics relative to our main results in Table 3, and thus may indicate comparatively poor model fit. Nonetheless, our results should be interpreted with some caution given these findings.

Residual Analysis Accounting for Correlations among SEL Outcomes

As discussed in the background section, the four SEL constructs we described are highly correlated. In particular, there is evidence that growth mindset and self-efficacy are associated with one another (West, Pier, et al., 2018). Thus, when these factors are taken in tandem, one might worry that there is evidence of an effect of reclassification on social awareness, for example, purely because the particular construct is also correlated with self-efficacy and growth mindset.

To test this theory, we perform a residual analysis to identify the effects of reclassification on each SEL construct after accounting for the other three (specifically, we regressed one SEL outcome on the other three, then used the residual from that model in our analyses). Figure S5 visualizes these residuals against binned binding scores. Similar to Figure S2, a small jump at the threshold is evident across all SEL constructs except social awareness.

To verify whether there are any jumps in SEL constructs in a causal framework while accounting for correlation within constructs, we repeat our main analyses using the residuals of SEL constructs as dependent variables, in lieu of the GPCM scaled scores used in the main results (though they are interpretable in the same units). Table 5 shows results from linear specifications, with and without district fixed effects. Akin to what was observed in the main results in Table 3, there is an increase in self-efficacy as a result of reclassification. Notably, however, the jump in self-efficacy is more modest in models using residuals. On one hand, this result could suggest that the effect or reclassification on self-efficacy alone is smaller than shown in other results. On the other, the estimate may decrease because, given the high correlation between self-efficacy and growth mindset, the true effect of self-efficacy is smaller once that correlation is accounted for (i.e., some of the true effect of self-efficacy is getting washed out by using residuals). The magnitude of the estimates for the other three SEL constructs also decreases when using residuals relative to what we observe in Table 3. These results suggest that, while the effect of reclassification on self-efficacy remains relatively robust to the use of residuals, estimates for the other three constructs may partially be an artifact of high correlations among SEL constructs, and with self-efficacy in particular. Nonetheless, as before, students who were reclassified experience higher levels of self-efficacy at the threshold.

Discussion

Extant literature indicates that reclassification into mainstream settings can be a critical moment in the academic trajectory of ELs. This study provides plausibly causal evidence that reclassification can influence students’ beliefs about their ability to succeed in schools. To our knowledge, it is the first to examine SEL skill development among this population at the precipice of reclassification, and therefore contributes to the current, limited array of evidence on the relationship between EL status and SEL outcomes (Niehaus & Adelson, 2014; West, Pier, et al., 2018). Reclassification policies, which have been highly decentralized and ad-hoc in nature for many decades (though are much more standardized under ESSA), could have different ramifications for students' SEL outcomes depending on whether the policy is aligned to student academic needs and readiness to forgo language supports.

Our results suggest that some aspect of exiting EL status has the potential to improve the way ELs perceive themselves as learners. Specifically, we find significant increases in self-efficacy post-reclassification that are robust to a range of sensitivity checks. Several theories could help explain this result. For example, if being an EL is a stigmatizing experience and reduces self-confidence, as described by Kanno and Kangas (2014), then one theory is that exiting EL status effectively removes that stigma. Alternatively, getting over the reclassification hurdle may bolster a student's belief that they are capable of successfully completing an academic task, the very definition of academic self-efficacy. Relatedly, if EL status restricts student access to academic content or confines them in settings where adults hold lower academic expectations, it is possible that reclassification allows students to access a higher quality academic environment and thus enables them to renew their belief in their academic success. While we are not able to isolate exactly which mechanism is driving the increase in self-efficacy (and there are certainly other theories that are plausible beyond what we have articulated), one might conjecture that more than one of these theories is at play in the results we observe.

Our findings have important implications for multiple stakeholders. For educators, if reclassification can positively affect self-efficacy in certain settings (as it appears to in ours), then educators may wish to investigate and address sources of lower self-efficacy among ELs who have not yet reclassified. For example, in the districts we studied, ELs who have not reclassified but are virtually identical to those who have in terms of their test scores still demonstrate lower self-efficacy compared to reclassified peers. Thus, educators may wish to identify SEL programs or interventions that can increase outcomes like self-efficacy among students who have not yet reclassified. This strategy may be particularly important given evidence that self-efficacy is associated with improved achievement and attainment among ELs (Soland & Sandilos, 2020; I. Umansky, 2018). In short, improving SEL outcomes may not only be beneficial in its own right; it may also improve academic outcomes, potentially even helping more students be academically prepared to reclassify. Additional research could determine whether students who experience increased academic self-efficacy as a result of reclassification from EL status also go on to improve in their academic achievement and attainment.

If the stigma of being labeled EL is what affects academic self-confidence in the settings we studied, those schools and districts can put more time and resources into increasing levels of inclusiveness and belonging among ELs, or implement other interventions that reduce the stigma. Regardless, further research on whether reclassification affects self-efficacy among ELs receiving different support service settings (e.g., bilingual education, sheltered instruction) can also help disentangle the mechanism through which SEL outcomes change for this student group, and how much they are affected by instructional approach. If there are differential impacts by EL instructional model, it would suggest that the transition to mainstream learning settings and/or a certain social-emotional aspect associated with a particular EL instructional setting contribute to students' academic confidence.

Our results also have implications for policymakers, though they are somewhat more complicated. Perhaps most importantly, our results show that reclassification *can* affect SEL-based outcomes, and policymakers (whether at the district, state, or federal level) should probably include SEL measures when examining effects of reclassification policies. However, there are two important caveats to this takeaway. First, the impact of reclassification policies—despite such policies being made more uniform under ESSA—are hugely contextual. For instance, the effects of reclassification could stem from removing academic supports, changing instructional settings, altering peer groups, or shifting teacher expectations (among other mechanisms). Therefore, results from our own study may not generalize to other settings since all of those mechanisms might differ locally—in fact, results probably will not generalize to most contexts. Although the complexity of reclassification and how much context can influence it likely impact generalizability, policymakers can nonetheless monitor how SEL is affected by reclassification in their specific contexts. Further, policymakers might do well to remember that reclassification is a potential mechanism for improving SEL as well as other downstream academic outcomes affected by constructs like self-efficacy (Almlund et al., 2011; Deming, 2017; Heckman et al., 2015; Heckman & Rubinstein, 2001).

Second, policymakers may look at our results and be tempted to conclude that ELs will benefit by having reclassification criteria lowered such that students become RFEP sooner and, potentially, see an earlier boost in constructs like self-efficacy. However, such an approach could have unintended consequences. Specifically, if students are reclassified before being ready academically, then less stringent criteria could undermine the very self-efficacy that such a policy was meant to improve. Therefore, in addition to keeping in mind contextual differences, policymakers should likely also carefully monitor how any changes to reclassification policies based in part on SEL findings impact academic outcomes. If nothing else, our results strongly suggest that SEL should be monitored when reclassifying ELs, and factored into decisions around criteria for reclassification.

For both educators and policymakers, our subgroup analyses may also provide useful information. We show that gains in self-efficacy upon reclassification are especially pronounced for female students and younger students. On average, students report much lower average self-efficacy in middle school compared to elementary school, and female students' self-efficacy declines more drastically than that of male students during these years (Fahle et al., 2019; West, Pier, et al., 2018). Our findings, meanwhile, suggest that the transition to RFEP status for ELs may be a way to mitigate declining self-efficacy, especially among females. Altogether, these subgroup analyses suggest that the effects of reclassification on SEL outcomes are likely not the same for all students under the EL umbrella.

Limitations

There are several key limitations to this study that should be considered, most related to generalizability. First, the data for this study came from a time when reclassification policies were non-standardized and varied widely in use (i.e., during the NCLB era). Thus, results will not necessarily generalize to students being reclassified in California today, as the assessments and criteria being used to determine reclassification thresholds have since changed, and the state has standardized guidelines for reclassification in more recent years. An extension of this research should seek to replicate this analysis using a more recent cohort of ELs who, in the ESSA era, were tested using more recent assessments and reclassified under more universal guidelines than the cohort observed in this study. Nonetheless, even under the consistency of procedures created by ESSA, states and districts maintain some latitude to reclassification criteria (such as the ability to choose which assessment to use to gauge English proficiency), suggesting policy decisions related to reclassification remain salient.

A second limitation pertaining to the data used is that results may not generalize to other states, nor even to other districts within California. For example, the districts examined in this study have a rather high concentration of EL students in their schools overall, a unique feature that can lead to differences in not only implementation of the reclassification policy, but instructional practice for ELs overall. A related limitation is that we do not have good data on the types of EL services provided in these districts. Thus, while we know that districts often used a combination of push-in and pull-out instructional methods, we cannot link a given student to a particular approach.

Similarly, as a result of the dearth of information on the services ELs in our sample receive, we can only speculate about possible mechanisms by which reclassification appears to affect outcomes like self-efficacy. For instance, we cannot conclude which aspect of the removal of the EL status (i.e., removal of stigma, broader access to academic content, or the increase of rigor in academic expectations) is the driving force in the observed increase in self-efficacy, as we are unable to differentiate ELs who receive certain services (e.g., bilingual versus sheltered instruction), or to quantify academic expectations held by teachers in classrooms of students, pre- and post-reclassification. Additionally, while we can assume that reclassification policies result in a change in instructional content for former ELs, we do not have information on the wide array of other changes that often occur upon reclassification in our administrative data, such as any changes in quality or composition of teachers, peers, or learning models; and in access to non EL-specific supports or programs students may have access to post-reclassification. These factors may also affect students upon reclassification. However, the decentralized nature of EL programming and lack of systematic data collection in these areas pose a challenge in further exploring the impact of these changes in our analysis, and there is little extant literature exploring this work. Case studies that investigate specific programs may be a better avenue for examining these changes upon reclassification.

Conclusion

The EL student population often has lower scores on academic achievement tests and attitudinal measures alike. Our study finds that, upon being reclassified into mainstream settings, academic self-efficacy among former EL students can improve significantly. These results suggest that some aspect of reclassification—whether it is the removal of the EL label, transitioning into an environment with higher academic expectations and standards, the combination of these factors, or some other mechanism entirely—can matter meaningfully for supporting ELs' beliefs and attitudes about academic success. Thus, educators and policymakers may benefit from monitoring SEL outcomes when implementing or changing reclassification policies, as well as better understanding which specific mechanisms can impact SEL constructs for ELs in a given context, both before and after reclassification.

References

Abedi, J. (2008). Classification system for English language learners: Issues and recommendations. *Educational Measurement: Issues and Practice*, *27*(3), 17–31. https://doi.org/10.1111/j.1745-3992.2008.00125.x

Almlund, M., Duckworth, A., Heckman, J., & Kautz, T. (2011). Personality Psychology and Economics. In E. A. Hanushek, S. Machin, & L. Woessman (Eds.), *Handbook of the Economics of Education* (Vol. 4, Issue 5500, pp. 1–181). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00001-8

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Betts, J., Hill, L., Bachofer, K., Hayes, J., Lee, A., & Zau, A. (2019). *English Learner Trajectories and Reclassification*.

Bolger, K. E., Patterson, C. J., Thompson, W. W., & Kupersmidt, J. B. (1995). Psychosocial Adjustment among Children Experiencing Persistent and Intermittent Family Economic Hardship. *Child Development*. https://doi.org/10.1111/j.1467-8624.1995.tb00926.x

Brooks, M. D. (2019). *Transforming Literacy Education for Long-Term English Learners*. Routledge. https://doi.org/10.4324/9781315151236

Bui, S. A. (2013). *How Do Limited English Proficient Students Affect Each Other's Educational Outcomes? Evidence from Student Panel Data*.

California Department of Education (2021). *Facts about English Learners in California.* Retrieved from https://www.cde.ca.gov/ds/ad/cefelfacts.asp

California English Language Development Test: Technical Report. (2015, November). Retrieved May 19, 2022, at https://www.cde.ca.gov/re/pr/documents/celdt1415techrpt.pdf

Callahan, R. M. (2013). *The English Learner Dropout Dilemma: Multiple Risks and Multiple Resources*.

Callahan, R. M., & Humphries, M. H. (2016). Undermatched? School-Based Linguistic Status, College Going, and the Immigrant Advantage. *American Educational Research Journal*, *53*(2), 263–295. https://doi.org/10.3102/0002831215627857

Callahan, R. M., Wilkinson, L., & Muller, C. (2010). Academic Achievement and Course Taking Among Language Minority Youth in U.S. Schools: Effects of ESL Placement. *Educational Evaluation and Policy Analysis*, *32*(1), 84–117. https://doi.org/10.3102/0162373709359805

Callahan, R. M., Wilkinson, L., Muller, C., & Frisco, M. (2009). ESL placement and schools: Effects on immigrant achievement. *Educational Policy*, *23*(2), 355–384. https://doi.org/10.1177/0895904807310034

CASEL. (2005). *Safe and Sound: An educational leader's guide to evidenced based SEL programs*.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression discontinuity designs. *Stata Journal, 17*, 372–404.

Catalano, R. F., Hawkins, J. D., Berglund, M. L., Pollard, J. A., & Arthur, M. W. (2002). Prevention science and positive youth development: Competitive or cooperative frameworks? *Journal of Adolescent Health*, *31*(6), 230–239. https://doi.org/10.1016/S1054-139X(02)00496-2

Choudhury, S., Blakemore, S. J., & Charman, T. (2006). Social cognitive development during adolescence. *Social Cognitive and Affective Neuroscience, 1*(3), 165–174. https://doi.org/10.1093/scan/nsl024

Cimpian, J. R., Thompson, K. D., & Makowski, M. B. (2017). Evaluating English Learner Reclassification Policy Effects Across Districts. *American Educational Research Journal*, *54*(1_suppl), 255S-278S. https://doi.org/10.3102/0002831216635796

Claro, S., & Loeb, S. (2019a). Self-Management Skills and Student Achievement Gains: Evidence from California's CORE Districts. In *PACE Working Paper*.

Claro, S., & Loeb, S. (2019b). *Students with Growth Mindset Learn More in School*.

Conley, D. T. (2008). Rethinking college readiness. *New Directions for Higher Education, 2008*(144), 3–13. https://doi.org/10.1002/he.321

Cunha, F., & Heckman, J. J. (2006). Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Journal of Human Resources*, *53*(4), 50–62. https://doi.org/10.1016/j.adolescence.2018.12.001

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*. https://doi.org/10.1080/10824669.2014.954044

de Boer, H., Timmermans, A. C., & van der Werf, M. P. C. (2018). The effects of teacher expectation interventions on teachers' expectations and student achievement: narrative review and meta-analysis. *Educational Research and Evaluation*, *24*(3–5), 180–200. https://doi.org/10.1080/13803611.2018.1550834

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, *132*(4), 1593–1640. https://doi.org/10.1093/qje/qjx022

Destin, M., Hanselman, P., Buontempo, J., Tipton, E., & Yeager, D. S. (2019). Do Student Mindsets Differ by Socioeconomic Status and Explain Disparities in Academic Achievement in the United States? *AERA Open*, *5*(3), 233285841985770. https://doi.org/10.1177/2332858419857706

Duckworth, A., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.92.6.1087

Duckworth, A., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2005.01641.x

Duckworth, A., Tsukayama, E., & May, H. (2010). Establishing Causality Using Longitudinal Hierarchical Linear Modeling: An Illustration Predicting Achievement From Self-Control. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550609359707

Dusenbury, L., Calin, S., Domitrovich, C., & Weissberg, R. P. (2015). *What Does Evidence-Based Instruction in Social and Emotional Learning Actually Look Like in Practice?* (Issue October).

Dweck, C. S. (2006). The Mindsets. In *Mindset: The New Psychology of Success*.

Dweck, C. S., Walton, G. M., & Cohen, G. L. (2011). *Academic tenacity: Mindset and skills that promote long-term learning*.

Education Commission of the States. (2020). *50-State Comparison: English Learner Policies*. https://www.ecs.org/50-state-comparison-english-learner-policies/

Equal Educational Opportunities Act, 20 U.S.C. § 1701 (1974).

Estrada, P., & Wang, H. (2018). Making English Learner Reclassification to Fluent English Proficient Attainable or Elusive: When Meeting Criteria Is and Is Not Enough. *American Educational Research Journal*, *55*(2), 207–242. https://doi.org/10.3102/0002831217733543

Fahle, E. M., Lee, M. G., & Loeb, S. (2019). *A Middle School Drop: Consistent Gender Differences in Students' Self-Efficacy*.

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*. https://doi.org/10.1353/esp.2016.0016

Flores, N., Kleyn, T., & Menken, K. (2015). Looking Holistically in a Climate of Partiality: Identities of Students Labeled Long-Term English Language Learners. *Journal of Language, Identity and Education*, *14*(2), 113–132. https://doi.org/10.1080/15348458.2015.1019787

Flores, S. M., & Drake, T. A. (2014). Does English language learner (ELL) identification predict college remediation designation? a comparison by race and ethnicity, and ELL waiver status. In *Review of Higher Education*. https://doi.org/10.1353/rhe.2014.0041

Frey, B. B. (2018). Every Student Succeeds Act. In *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. https://doi.org/10.4135/9781506326139.n245

Gándara, P., Maxwell-Jolly, J., & Driscoll, A. (2005). *Listening to teachers of English Language Learners: A survey of California teachers' challenges, experiences, and professional development needs*.

Gaspar, Tania, Cerqueira, A., Branquinho, C., & Matos, M. G. De. (2018). *Original Research Article Original Research Article Open Access Dimensions of Social and Personal Skills in Children and Adolescents: Age and Gender Differences*. *08*(2), 18394–18400.

Gelman, A., & Imbens, G. (2019). Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. *Journal of Business and Economic Statistics*, *37*(3), 447–456. https://doi.org/10.1080/07350015.2017.1366909

Goldenberg, C., & Coleman, R. (2010). *Promoting Academic Achievement among English Learners: A Guide to the Research*. Corwin.

González, N., Moll, L. C., & Amanti, C. (2006). *Funds of knowledge: Theorizing practices in households, communities, and classrooms*. Routledge.

Heckman, J. J., Humphries, J. E., Kautz, T., Heckman, J. J., & Kautz, T. (2015). Achievement Tests and the Role of Character in American Life. In *The Myth of Achievement Tests*. https://doi.org/10.7208/chicago/9780226100128.003.0001

Heckman, J. J., & Kautz, T. (2013). Fostering and Measuring Skills: Interventions that Improve Character and Cognition. In *National Bureau of Economic Research* (Vol. 1). https://doi.org/10.1017/CBO9781107415324.004

Heckman, J. J., & Rubinstein, Y. (2001). The benefits of skill: The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, *91*(2), 145–154.

Hill, L. E., Weston, M., & Hayes, J. M. (2014). Reclassification of English Learner Students in California. *Public Policy Institute of California*.

Honigsfeld, A., & Dove, M. G. (2019). *Collaborating for English Learners: A Foundational Guide to Integrated Practices* (2nd ed.).

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education*, *28*(1), 1–35. https://doi.org/10.1007/s10212-011-0097-y

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, *79*(3), 933–959. https://doi.org/10.1093/restud/rdr043

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2007.05.001

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in Children's Self-Competence and Values: Gender and Domain Differences across Grades One through Twelve. *Child Development*, *73*(2), 509–527.

Johnson, A. (2020). The Impact of English Learner Reclassification on High School Reading and Academic Progress. *Educational Evaluation and Policy Analysis*, *42*(1), 46–65. https://doi.org/10.3102/0162373719877197

Kanno, Y., & Kangas, S. E. N. (2014). "I'm Not Going to Be, Like, for the AP": English Language Learners' Limited Access to Advanced College-Preparatory Courses in High School. *American Educational Research Journal*, *51*(5), 848–878. https://doi.org/10.3102/0002831214544716

LeClair, C., Doll, B., Osborn, A., & Jones, K. (2009). English language learners' and non–English language learners' perceptions of the classroom environment. *Psychology in the Schools*, *46*: 568-577. https://doi.org/10.1002/pits.20398

Lee, D. S., & Lemieux, T. (2010). Regression Discontinuity designs in economics. *Journal of Economic Literature*. https://doi.org/10.1257/jel.48.2.281

Mavrogordato, M., & White, R. S. (2017). Reclassification Variation: How Policy Implementation Guides the Process of Exiting Students From English Learner Status. *Educational Evaluation and Policy Analysis*, *39*(2), 281–310. https://doi.org/10.3102/0162373716687075

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*(2):698–714.

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01398-0

Merolla, D. M. (2017). Self-efficacy and Academic Achievement. *Sociological Perspectives, 60*(2), 378–393. https://doi.org/10.1177/0731121416629993

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*.

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*. https://doi.org/10.1177/014662169201600206

National Center for Education Statistics. (2019). *English Language Learners in Public Schools*. https://nces.ed.gov/programs/coe/indicator_cgf.asp

Niehaus, K., & Adelson, J. L. (2014). School Support, Parental Involvement, and Academic and Social-Emotional Outcomes for English Language Learners. *American Educational Research Journal*. https://doi.org/10.3102/0002831214531323

Oakes, J. (1985). *Keeping track: How schools structure inequality*. Yale University Press.

Pajares, F., Britner, S. L., & Valiante, G. (2000). Relation between Achievement Goals and Self-Beliefs of Middle School Students in Writing and Science. *Contemporary Educational Psychology*. https://doi.org/10.1006/ceps.1999.1027

Papay, J. P., Willett, J. B., & Murnane, R. J. (2011). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics*. https://doi.org/10.1016/j.jeconom.2010.12.008

Papini, D. R., Farmer, F. F., Clark, S. M., Micka, J. C., & Barnett, J. K. (1990). Early adolescent age and gender differences in patterns of emotional self-disclosure to parents and friends. In *Adolescence* (Vol. 25, Issue 100, pp. 959–976). Libra Publishers.

Parrish, T. B., Perez, M., Merickel, A., & Linquanti, R. (2006). *Effects of the Implementation of Proposition 227 on the Education of English Learners, K-12.*

Pintrich, P. R., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In *Development of achievement motivation.* (pp. 249–284). Academic Press. https://doi.org/10.1016/B978-012750053-9/50012-7

Pope, N. G. (2016). The Marginal effect of K-12 English language development programs: Evidence from Los Angeles Schools. *Economics of Education Review*, *53*, 311–328. https://doi.org/10.1016/j.econedurev.2016.04.009

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher Order Instructional Goals in Secondary Schools: Class, Teacher, and School Influences. *American Educational Research Journal*, *30*(3), 523–553. https://doi.org/10.3102/00028312030003523

Reardon, S. F., & Galindo, C. (2009). The Hispanic-White achievement gap in math and reading in the elementary grades. *American Educational Research Journal*, *46*(3), 853–891. https://doi.org/10.3102/0002831209333184

Reardon, S. F., & Robinson, J. P. (2012). Regression Discontinuity Designs With Multiple Rating-Score Variables. *Journal of Research on Educational Effectiveness*. https://doi.org/10.1080/19345747.2011.609583

Robinson, J. P. (2011). Evaluating Criteria for English Learner Reclassification: A Causal-Effects Approach Using a Binding-Score Regression Discontinuity Design With Instrumental Variables. *Educational Evaluation and Policy Analysis*, *33*(3), 267–292. https://doi.org/10.3102/0162373711407912

Robinson-Cimpian, J. P., & Thompson, K. D. (2016). The Effects of Changing Test-Based Policies for Reclassifying English Learners. *Journal of Policy Analysis and Management*, *35*(2), 279–305. https://doi.org/https://doi.org/10.1002/pam.21882

Robinson-Cimpian, J. P., Thompson, K. D., & Umansky, I. (2016). Research and Policy Considerations for English Learner Equity. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 129–137. https://doi.org/10.1177/2372732215623553

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*. https://doi.org/10.1007/BF02322211

Rumberger, R. W., & Gándara, P. (2004). Seeking equity in the education of California's English learners. In *Teachers College Record*. https://doi.org/10.1111/j.1467-9620.2004.00426.x

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*. https://doi.org/10.1007/bf03372160

Santibañez, L., & Umansky, I. (2018). English Learners: Charting Their Experiences and Mapping Their Futures in California Schools. In *Getting Down to Facts II* (Issue September).

Schunk, D. H., & Meece, J. L. (2006). Self-Efficacy Development in Adolescents. *Self-Efficacy Beliefs of Adolescents*, 71–96.

Schunk, D. H., & Pajares, F. (2002). The Development of Academic Self-Efficacy. *Development of Achievement Motivation*, *1446*, 15–31. https://doi.org/10.1016/B978-012750053-9/50003-6

Shin, N. (2017). The Effects of the Initial English Language Learner Classification on Students' Later Academic Outcomes. *Educational Evaluation and Policy Analysis*, *XX*(X), 016237371773737. https://doi.org/10.3102/0162373717737378

Sklad, M., Diekstra, R., de Ritter, M., Ben, J., & Gravesteijn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools*, *49*(9), 892–909. https://doi.org/10.1002/pits

Slama, R. B. (2014). Investigating Whether and When English Learners Are Reclassified Into Mainstream Classrooms in the United States: A Discrete-Time Survival Analysis. *American Educational Research Journal*, *51*(2), 220–252. https://doi.org/10.3102/0002831214528277

Snipes, J., & Tran, L. (2017). Growth Mindset, Performance Avoidance, and Academic Behaviors in Clark County School District (REL 2017-226). In *Regional Educational Laboratory West* (Issue April).

Soland, J. (2022). Evidence That Selecting an Appropriate Item Response Theory–Based Approach to Scoring Surveys Can Help Avoid Biased Treatment Effect Estimates. *Educational and Psychological Measurement*, *82*(2), 376–403. https://doi.org/10.1177/00131644211007551

Soland, J., Hamilton, L. S., & Stecher, B. M. (2013). Measuring 21st century competencies: Guidance for educators. In *Asia Society Global Cities Education Network Report*.

Soland, J., Kuhfeld, M., Wolk, E., & Bi, S. (2019). Examining the State-Trait Composition of Social-Emotional Learning Constructs: Implications for Practice, Policy, and Evaluation.

Journal of Research on Educational Effectiveness, *12*(3), 550–577. https://doi.org/10.1080/19345747.2019.1615158

Soland, J., Rimm-Kaufman, S. E., Kuhfeld, M., & Ventura-Abbas, N. (2022). Empirical benchmarks for changes in social and emotional skills over time. *Child Development*. https://doi.org/10.1111/cdev.13741

Soland, J., & Sandilos, L. E. (2020). English Language Learners, Self-efficacy, and the Achievement Gap: Understanding the Relationship between Academic and Social-Emotional Growth. *Journal of Education for Students Placed at Risk (JESPAR)*, 1–25. https://doi.org/10.1080/10824669.2020.1787171

Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of Dual-Language Immersion Programs on Student Achievement: Evidence From Lottery Data. *American Educational Research Journal*, *54*(1), 282S-306S. https://doi.org/10.3102/0002831216634463

Thompson, K. D. (2015). Questioning the Long-Term English Learner Label: How Categorization Can Blind Us to Students' Abilities. *Teachers College Record*, *117*, 50.

Umansky, I. (2016a). Leveled and Exclusionary Tracking. *American Educational Research Journal*, *53*(6), 1792–1833. https://doi.org/10.3102/0002831216675404

Umansky, I. (2016b). To Be or Not to Be EL: An Examination of the Impact of Classifying Students as English Learners. *Educational Evaluation and Policy Analysis*, *38*(4), 714–737. https://doi.org/10.3102/0162373716664802

Umansky, I. (2018). *State Policies to Advance English Learners' Experiences and Outcomes in California's Schools* (Issue September).

Umansky, I., & Dumont, H. (2021). English Learner Labeling: How English Learner Classification in Kindergarten Shapes Teacher Perceptions of Student Skills and the Moderating Role of Bilingual Instructional Settings. *American Educational Research Journal*, *58*(5), 993–1031. https://doi.org/10.3102/0002831221997571

Umansky, I., & Reardon, S. (2014). Reclassification Patterns Among Latino English Learner Students in Bilingual, Dual Immersion, and English Immersion Classrooms. *American Educational Research Journal*, *51*(5), 879–912. https://doi.org/10.3102/0002831214545110

Umansky, I., Reardon, S. F., Hakuta, K., Thompson, K. D., Estrada, P., Hayes, K., Maldonado, H., Tandberg, S., & Goldenberg, C. (2015). Improving the opportunities and outcomes of California's students learning English: Findings from school district-university collaborative partnerships. *PACE Policy Brief*, *15–1*(October), 1–16. https://doi.org/10.1016/S0264-410X(12)01439-9

Valdés, G. (1998). The world outside and inside schools: Language and immigrant children. *Educational Researcher*, *27*(6), 4–18. https://doi.org/10.3102/0013189X027006004

Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of Four Instructional Programs Designed to Serve English Learners: Variation by Ethnicity and Initial English Proficiency. *Educational Evaluation and Policy Analysis*, *37*(4), 612–637. https://doi.org/10.3102/0162373715573310

Valenzuela, A. (1999). *Subtractive Schooling: U.S.-Mexican Youth and the Politics of Caring*. State University of New York Press.

Villegas, L., & Pompa, D. (2020). *The Patchy Landscape of State English Learner Policies under ESSA.*

Weissberg, R. P., & Cascarino, J. (2013). Academic learning + social-emotional learning = national priority. *Phi Delta Kappan*, *95*(2), 8–13. https://doi.org/10.1177/003172171309500203

West, M. R., Buckley, K., Krachman, S. B., & Bookman, N. (2018). Development and implementation of student social-emotional surveys in the CORE Districts. *Journal of Applied Developmental Psychology*. https://doi.org/10.1016/j.appdev.2017.06.001

West, M. R., Pier, L., Fricke, H., Hough, H., Loeb, S., Meyer, R. H., & Rice, A. B. (2020). Trends in Student Social-Emotional Learning: Evidence From the First Large-Scale Panel Student Survey. *Educational Evaluation and Policy Analysis, 42*(2), 279–303. https://doi.org/10.3102/0162373720912236

Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children’s domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, *27*(4), 552–565. https://doi.org/10.1037/0012-1649.27.4.552

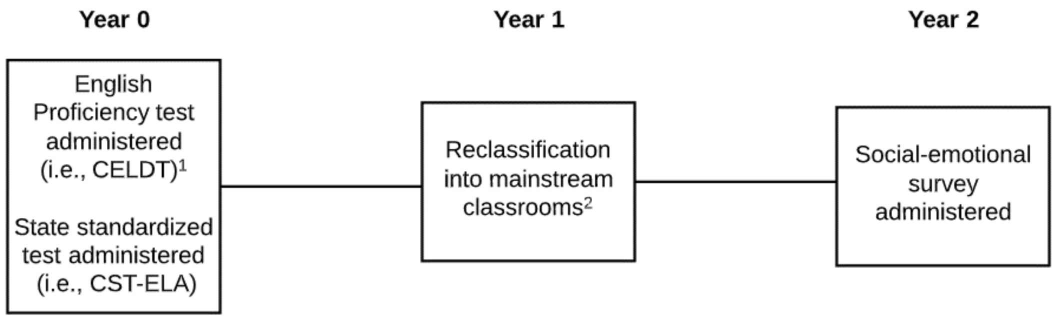
Wigfield, A., Eccles, J. S., & Pintrich, P. R. (1996). Development between the ages of 11 and 25. In *Handbook of educational psychology.* (pp. 148–185). Prentice Hall International.

Yoon, B. (2008). Uninvited guests: The influence of teachers’ roles and pedagogies on the positioning of English language learners in the regular classroom. *American Educational Research Journal*, *45*(2), 495–522. https://doi.org/10.3102/0002831208316200

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*. https://doi.org/10.3102/00028312029003663

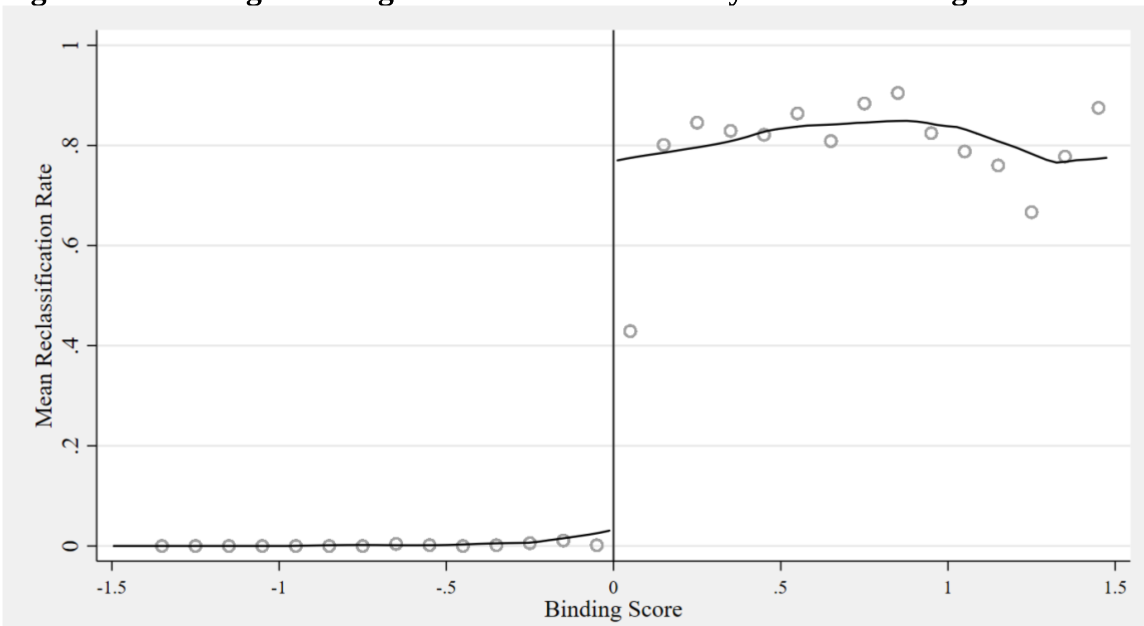
Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, *82*(1), 51–59. https://doi.org/10.1037/0022-0663.82.1.51

Figure 1. Timeline for Reclassification and Outcome Observation.



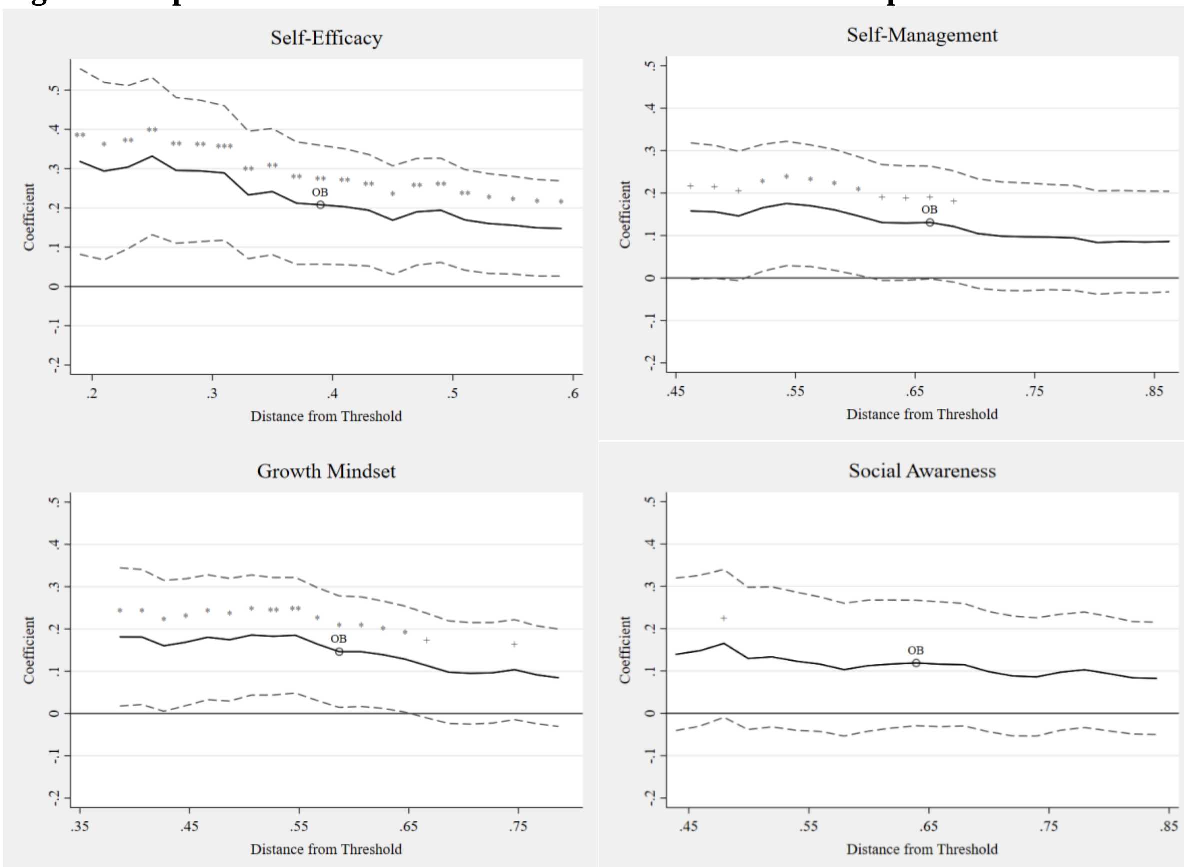
[1] A proportion of the sample has two CELDT scores: One from Year 0 and a second from Year 1.
[2] Reclassification only granted to ELs who are eligible based on test-based and qualitative criteria per district guidelines.

Figure 2. First stage: Average Reclassification Rate by Binned Binding Scores



Notes: The binding score is the student's lowest standardized score on the CST-ELA, overall CELDT, and subsections of the CELDT, centered at the cut point. The vertical axis is the average reclassification rate in each bin. Bin width = 0.10 SD.

Figure 3. Impacts of Reclassification on SEL Constructs at Multiple Bandwidths



Notes: The solid and dashed lines represent beta coefficients and 95% confidence intervals, respectively, of preferred model results using 20 alternative samples, in addition to the sample determined by the optimal bandwidth (OB). Coefficient from OB sample labeled on figure. OB sample contains students with binding scores no greater than 0.39 SD above and below the threshold. Additional alternative samples are determined by adding and subtracting intervals of 0.02 SD from the OB binding score threshold. Stars denote significance. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 1. SEL Constructs: Summarized Definitions and Sample Items

Skill	Definition	Sample item
Self-management	The ability to regulate one's emotions, thoughts, and behaviors effectively in different situations. This includes managing stress, delaying gratification, motivating oneself, and selecting and working toward personal and academic goals.	Please answer how often you did the following during the last 30 days. During the past 30 days... I remained calm even when criticized or otherwise provoked.
Growth mindset	The belief that one's abilities grow with effort. Students with a growth mindset see effort as necessary for success, embrace challenges, learn from criticism, and persist in the face of setbacks.	Please indicate how true each of the following statements is for you: My intelligence is something that I can't change very much.
Self-efficacy	The belief in one's ability to succeed in achieving an outcome or reaching a goal. Self-efficacy reflects confidence in the ability to exert control over one's motivation, behavior, and environment.	How confident are you about the following at school? I can do well on my tests even when they are difficult.
Social awareness	The ability to take the perspective of and empathize with others from diverse backgrounds and cultures, to understand social and ethical norms for behavior, and to recognize family, school, and community resources and supports.	During the past 30 days... How carefully did you listen to other people's points of view?

Note: Definitions of SEL constructs from West, Buckley, Krachman, & Bookman (2018). Items are derived from SEL surveys administered to students.

Table 2. Descriptive Statistics of Analytic Sample and Subsamples by District.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	All Districts		District 1		District 2		District 3	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Female	0.49	(0.50)	0.49	(0.50)	0.48	(0.50)	0.49	(0.50)
Race/Ethnicity								
Latinx	0.90	(0.29)	0.80	(0.40)	0.90	(0.30)	0.98	(0.13)
Asian	0.09	(0.28)	0.19	(0.39)	0.09	(0.29)	0.01	(0.12)
Other/Multi Race	0.01	(0.08)	0.01	(0.11)	0.01	(0.08)	0.00	(0.05)
Free/Reduced Lunch	0.41	(0.49)	0.40	(0.49)	0.15	(0.35)	0.63	(0.48)
Reclassification Rate	0.25	(0.43)	0.22	(0.41)	0.16	(0.37)	0.35	(0.48)
Grade Level at Reclassification								
Grade 4	0.32	(0.47)	0.25	(0.44)	0.29	(0.45)	0.38	(0.49)
Grade 5	0.23	(0.42)	0.25	(0.44)	0.23	(0.42)	0.21	(0.41)
Grade 6	0.18	(0.38)	0.18	(0.38)	0.19	(0.39)	0.17	(0.38)
Grade 7	0.15	(0.36)	0.18	(0.39)	0.16	(0.37)	0.12	(0.32)
Grade 8	0.12	(0.33)	0.13	(0.33)	0.13	(0.34)	0.12	(0.32)
English Proficiency Test (CELDT)								
Overall	-0.31	(0.69)	-0.48	(0.66)	-0.30	(0.65)	-0.21	(0.71)
Listening	0.64	(0.83)	0.54	(0.82)	0.60	(0.80)	0.75	(0.85)
Speaking	0.92	(0.77)	0.86	(0.79)	0.99	(0.77)	0.90	(0.74)
Reading	0.27	(0.76)	0.05	(0.76)	0.34	(0.75)	0.37	(0.74)
Writing	0.57	(0.70)	0.38	(0.70)	0.56	(0.65)	0.72	(0.71)
State Standardized Test (CST-ELA)	-0.21	(0.67)	-0.35	(0.64)	-0.13	(0.66)	-0.16	(0.69)
SEL Skills								
Growth Mindset	-0.43	(1.00)	-0.24	(1.18)	-0.49	(0.98)	-0.53	(0.85)
Social Awareness	0.03	(1.17)	0.06	(1.22)	0.13	(1.20)	-0.07	(1.10)
Self-Efficacy	-0.17	(0.93)	-0.19	(0.96)	-0.11	(0.92)	-0.21	(0.91)
Self-Management	-0.11	(1.07)	-0.00	(1.16)	-0.03	(1.06)	-0.24	(1.00)
Number of Observations	10637		3145		3231		4261	

The statistics above pertains to a sample of students whose binding scores are within 1.5 standard deviations of the cut point of 0. All test scores are standardized and centered on the cutoff. SEL skills are IRT GPCM scaled scores centered at zero and are measured post-reclassification. Reclassification rate is generated by dividing the number of students reclassified in Year 1 by the total number of English Learners in Year 0.

Table 3. Main Results: Impact of Reclassification on SEL Outcomes

	Linear				Quadratic			
	(1)		(2)		(3)		(4)	
Self-Efficacy	0.195	**	0.208	**	0.343	**	0.352	**
	(0.076)		(0.077)		(0.128)		(0.128)	
N	4851		4851		4851		4851	
F	1309		1514		437		514	
Self-Management	0.104		0.131	+	0.139		0.159	
	(0.067)		(0.068)		(0.108)		(0.109)	
N	7194		7194		7194		7194	
F	2410		2744		863		1001	
Growth Mindset	0.131	*	0.146	*	0.223	*	0.232	*
	(0.067)		(0.067)		(0.106)		(0.106)	
N	6685		6685		6685		6685	
F	2159		2457		784		919	
Social Awareness	0.094		0.119		0.125		0.138	
	(0.074)		(0.076)		(0.121)		(0.122)	
N	7069		7069		7069		7069	
F	2354		2664		831		978	
District FE	No		Yes		No		Yes	

Note: Robust standard errors in parentheses. Each coefficient represents a separate 2SLS regression. F-statistic reported from the first stage. Sample sizes (N) are the same between the first and second stage. Other coefficients omitted from display. All coefficients report impact at the optimal bandwidth, which ranges between .32 and .77 for the four outcomes.

+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Table 4. Heterogeneity Analysis of Main Results: Impact of Reclassification on SEL Outcomes by Subgroup

	Gender				Socioeconomic Status				Grade Level		
	(1)		(2)		(3)		(4)		(8)		(9)
	Male		Female		Non-FRPL		FRPL		Elementary		Middle
Self-Efficacy	0.083		0.217	*	0.186	+	0.187	+	0.196	*	0.106
	(0.081)		(0.097)		(0.098)		(0.100)		(0.090)		(0.109)
N	3813		2724		3629		1898		2755		2950
F	1288		991		873		1136		1691		587
Self-Management	0.107		0.233	*	0.115		0.08		0.151	+	0.127
	(0.108)		(0.112)		(0.102)		(0.087)		(0.091)		(0.148)
N	3028		2851		4163		3127		3440		2459
F	959		1060		1071		2285		2224		436
Growth Mindset	0.26	*	0.138		0.167	+	0.102		0.233	**	0.101
	(0.123)		(0.090)		(0.094)		(0.097)		(0.088)		(0.131)
N	2476		3289		4331		2183		3379		2567
F	744		1320		1162		1413		2165		466
Social Awareness	0.025		0.226	+	0.107		0.071		0.138		0.132
	(0.103)		(0.126)		(0.124)		(0.076)		(0.122)		(0.169)
N	3710		2814		3678		4389		2479		2389
F	1234		1042		891		3643		1605		413

Notes: Robust standard errors in parentheses. Each coefficient represents a separate 2SLS regression. F-statistics reported from the first stage. Sample sizes (N) are the same between the first and second stage. Other coefficients omitted from display. All coefficients report impact at the optimal bandwidth.
+ p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table 5. Replication of Main Model Using Residuals as Dependent Variables.

	(1)		(2)	
Self-Efficacy	0.118	+	0.121	+
	(0.063)		(0.064)	
N	4880		4880	
F	1327		1531	
Self-Management	0.058		0.074	
	(0.058)		(0.059)	
N	6525		6525	
F	2065		2353	
Growth Mindset	0.079		0.086	
	(0.060)		(0.061)	
N	7312		7312	
F	2509		2850	
Social Awareness	-0.014		-0.009	
	(0.065)		(0.066)	
N	6138		6138	
F	1902		2176	
District FE	No		Yes	

Note: Robust standard errors in parentheses. Each coefficient represents a separate 2SLS linear regression. F-statistic reported from the first stage. Sample sizes (N) are the same between the first and second stage. Other coefficients omitted from display. All coefficients report impact at the optimal bandwidth, which ranges between .40 and .68. Each outcome is a residual produced from linear regressions predicting the construct in question using the remaining three constructs.

+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Supplementary Online Materials for

Does Reclassification Change How English Learners Feel about School and Themselves? Evidence from a Regression Discontinuity Design

Contents

Evidence on SEL Constructs

Self-Efficacy

Self-efficacy refers to how individuals judge their own abilities to perform certain tasks or actions (Bandura, 1982a, 1997). Bandura (1982b, 1997) argues 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) shows 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 and educational attainment (Zimmerman et al., 1992).Much of the literature supports the theory that fostering self-efficacy skills results in an increase in cognitive skills via test scores or assignment/course grades. 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 et al., 2010). Researchers have also shown that academic self-efficacy, one of the SEL constructs measured in this study, is higher among White students relative to non-White students (Kitsantas et al., 2011; Pajares & Kranzler, 1995) and that SEL gaps are particularly evident between students of high- and low-poverty backgrounds (Bradley & Corwyn, 2002; Brooks-Gunn & Duncan, 1997; West, Pier, et al., 2018).

Self-Management

Self-management is defined by the districts in our study as the way a student maintains control over his or her thoughts, behaviors, and emotions. Self-management is an umbrella term that captures aspects of self-regulated learning like the ability to stay focused (Pintrich & de Groot, 1990), as well as classroom behaviors indicative of self-regulation like coming to class prepared (Farrington et al., 2012). Examples of questions in this construct include asking students if they can work independently, follow directions, and are able to stay focused on their schoolwork.

Research suggests that self-regulation is a direct function of how motivated a student is (Pintrich, 1999; Schunk, 2005). As a result, self-management has been shown to be positively associated with higher report card grades and GPAs (Duckworth & Seligman, 2005; Lee Duckworth et al., 2010). A recent study has shown that self-management is a better predictor of academic growth relative to other measures of SEL (Claro & Loeb, 2019a).

Growth Mindset

Growth mindset refers to the ways that students view their intelligence on a continuum from fixed to malleable (Blackwell, Trzesniewski, & Dweck, 2007; Dweck, 2006). Students with a fixed mindset believe intelligence is a static trait: they have a certain amount of intelligence that cannot be changed (Dweck, 2006). By contrast, students with a growth mindset believe they can change their intelligence over time. Growth mindset can have a profound impact on whether students exert effort, especially in schooling contexts (Dweck, 2010). As a result, students randomized to receive growth mindset interventions demonstrated higher GPAs (Aronson, Fried, & Good, 2002), which are strongly predictive of finishing high school (Allensworth, 2013).

Evidence on the relationship between growth mindset and academic outcome has been mixed. Many interventions, report small, positive results in achievement from mindset interventions (Aronson et al., 2002; Blackwell et al., 2007; Paunesku et al., 2015; Yeager et al., 2016). A recent study using the same growth mindset measures as the present study also found meaningful growth in English and math corresponding to students with growth mindsets, relative to fixed mindset (Claro & Loeb, 2019b). In contrast, a meta-analysis of mindset as a moderator and mediator of academic achievement found that the effects are generally weak, although interventions meaningfully increased achievement among at-risk subgroups such as low-income and low-achieving students (Sisk et al., 2018).

Social Awareness

Social awareness can be defined in different ways and oftentimes refers to a cluster of constructs. In the context of this study, social awareness is primarily defined as “the ability to understand social norms for behavior” (CASEL, 2005). The relationship between social awareness and achievement and attainment have been described as “tenuous” (Farrington et al., 2012, p. 11). Nonetheless, there are indirect ways that social awareness impacts learning and attainment (Farrington et al., 2012). For example, students with less social awareness are likelier to disrupt their classes, which can reduce the quality of classroom instruction and make it harder for students to engage with the material (Epstein et al., 2008).

Complete SEL Survey Questionnaire

Self-Management

First, we'd like to learn more about your behavior, experiences, and attitudes related to school. Please answer how often you did the following during the past 30 days. During the past 30 days…

1. I came to class prepared.
2. I remembered and followed directions.
3. I got my work done right away instead of waiting until the last minute.
4. I paid attention, even when there were distractions.
5. I worked independently with focus.
6. I stayed calm even when others bothered or criticized me.
7. I allowed others to speak without interruption.
8. I was polite to adults and peers.
9. I kept my temper in check.

(Almost Never, Once in a While, Sometimes, Often, Almost All the Time)

Growth Mindset

In this section, please think about your learning in general. Please indicate how true each of the following statements is for you:

10. My intelligence is something that I can't change very much.
11. Challenging myself won't make me any smarter.
12. There are some things I am not capable of learning.
13. If I am not naturally smart in a subject, I will never do well in it.

(Not at All True, A Little True, Somewhat True, Mostly True, Completely True)

Self-efficacy

How confident are you about the following at school?

14. I can earn an A in my classes.

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

16. I can master the hardest topics in my classes.

17. I can meet all the learning goals my teachers set.

(Not at All Confident, A Little Confident, Somewhat Confident, Mostly Confident, Completely Confident)

Social Awareness

In this section, please help us better understand your thoughts and actions when you are with other people. Please answer how often you did the following during the past 30 days. During the past 30 days…

18. How carefully did you listen to other people's points of view?

(Not Carefully at All, Slightly Carefully, Somewhat Carefully, Quite Carefully, Extremely Carefully)

19. How much did you care about other people's feelings?

(Did Not Care at All, Cared a Little Bit, Cared Somewhat, Cared Quite a Bit, Cared a Tremendous Amount)

20. How often did you compliment others' accomplishments?

(Almost Never, Once in a While, Sometimes, Often, Almost All the Time)

21. How well did you get along with students who are different from you?

(Did Not Get Along at All, Got Along a Little Bit, Got Along Somewhat, Got Along Pretty Well, Got Along Extremely Well)

22. How clearly were you able to describe your feelings?

(Not at All Clearly, Slightly Clearly, Somewhat Clearly, Quite Clearly, Extremely Clearly)

23. When others disagreed with you, how respectful were you of their views?

(Not at All Respectful, Slightly Respectful, Somewhat Respectful, Quite Respectful, Extremely Respectful)

24. To what extent were you able to stand up for yourself without putting others down?

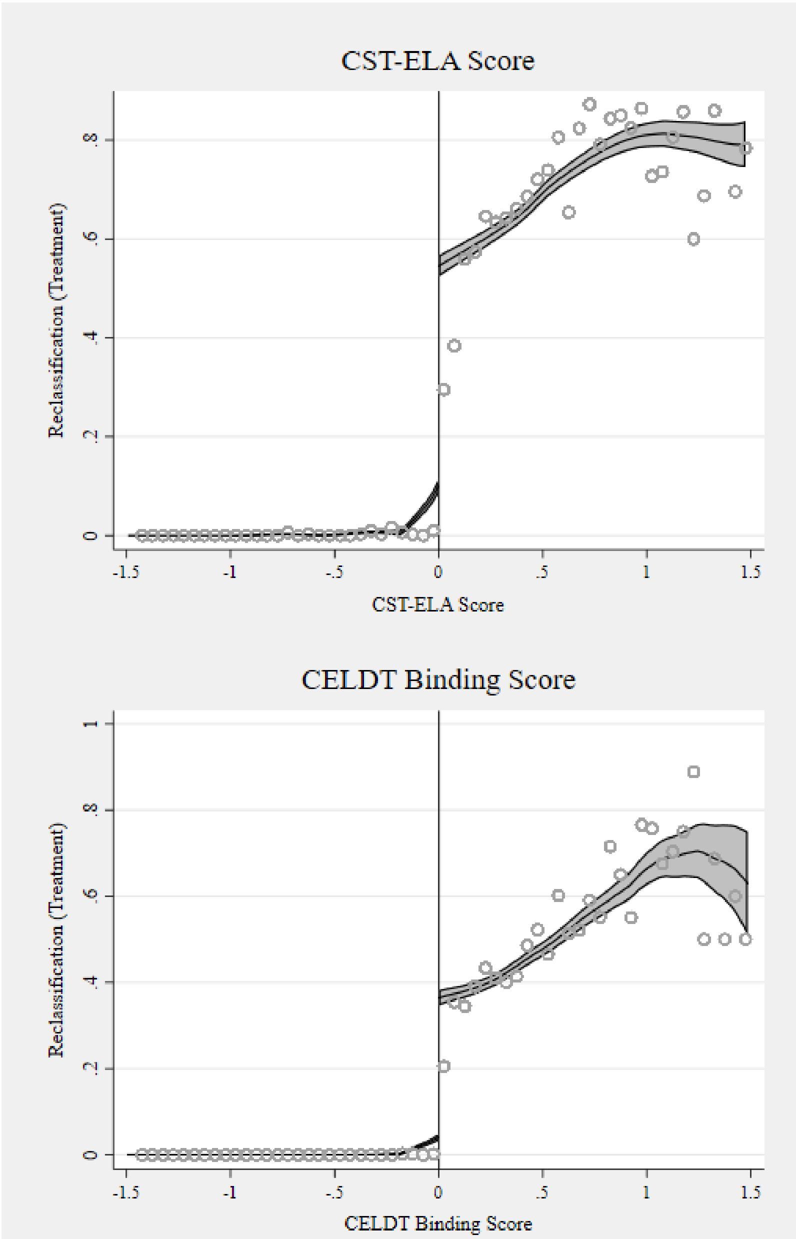
(Not at All, A Little Bit, Somewhat, Quite a Bit, A Tremendous Amount)

25. To what extent were you able to disagree with others without starting an argument?

(Not at All, A Little Bit, Somewhat, Quite a Bit, A Tremendous Amount)

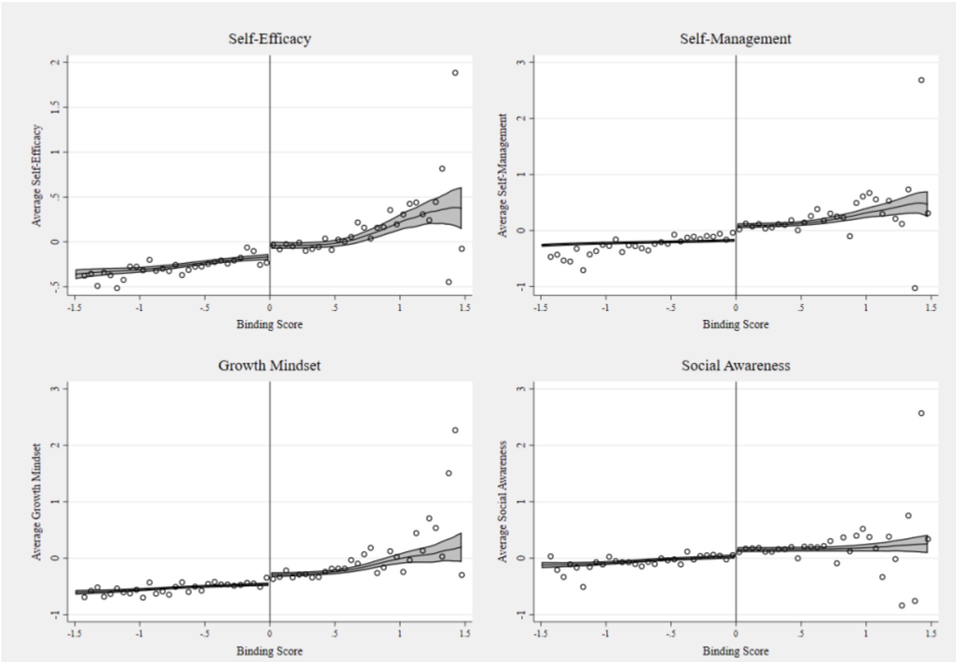
Supplementary Results

Figure S1. First Stage Visualization of Treatment Status by Assessment Used in Binding Score.



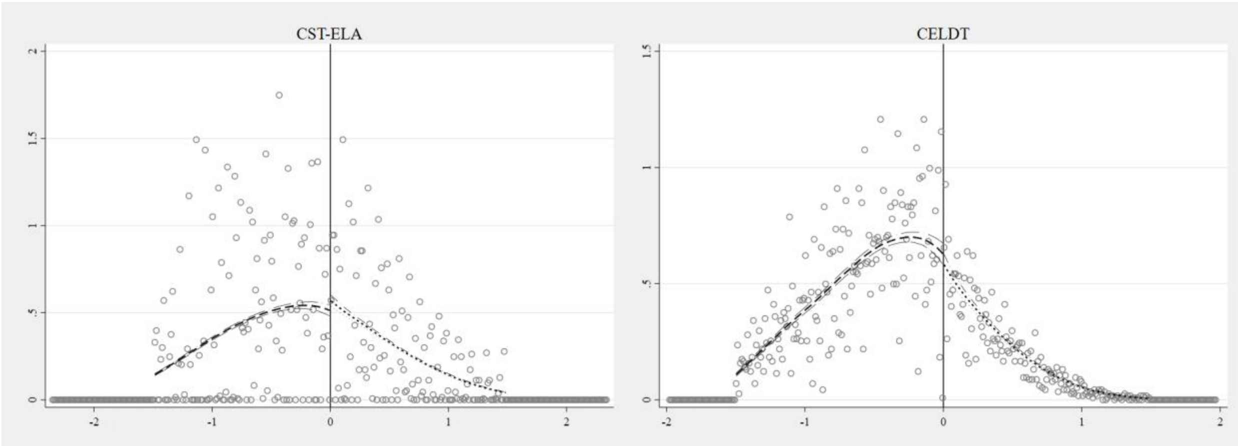
Notes: The horizontal axis of the top and bottom panels are, respectively: standardized CST-ELA centered at the cut point; and binding score of the CELDT (i.e., standardized minimum value of the overall CELDT score and four subsections of the CELDT – reading, writing, speaking, and listening – centered at the cut point). The vertical axis is the average reclassification rate pertaining to each bin. Bin width = 0.05 SD. Gray areas are 95% CI.

Figure S2: Bin Plots of SEL Constructs by Binding Score.



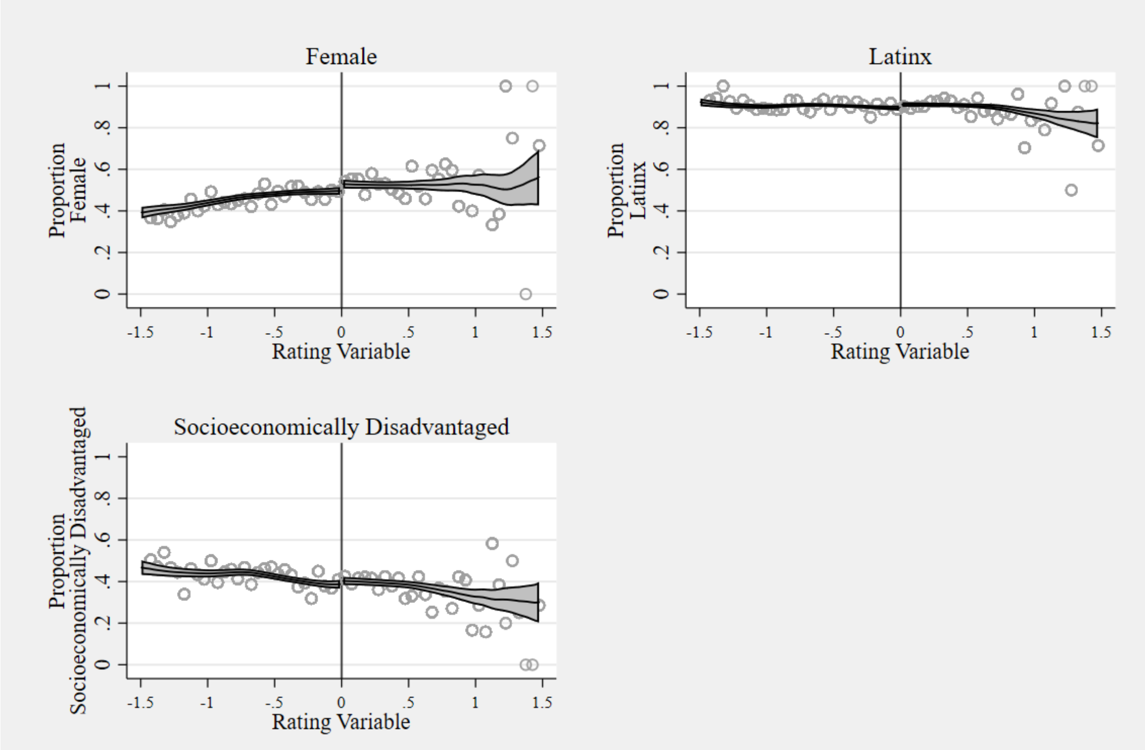
Notes: Each point pertains to the average value of each SEL construct for students with binding scores falling within each bin. Bin width = 0.05 SD. Gray areas are 95% CI.

Figure S3: McCrary Density Test: Distribution of Observations by Binding Score.



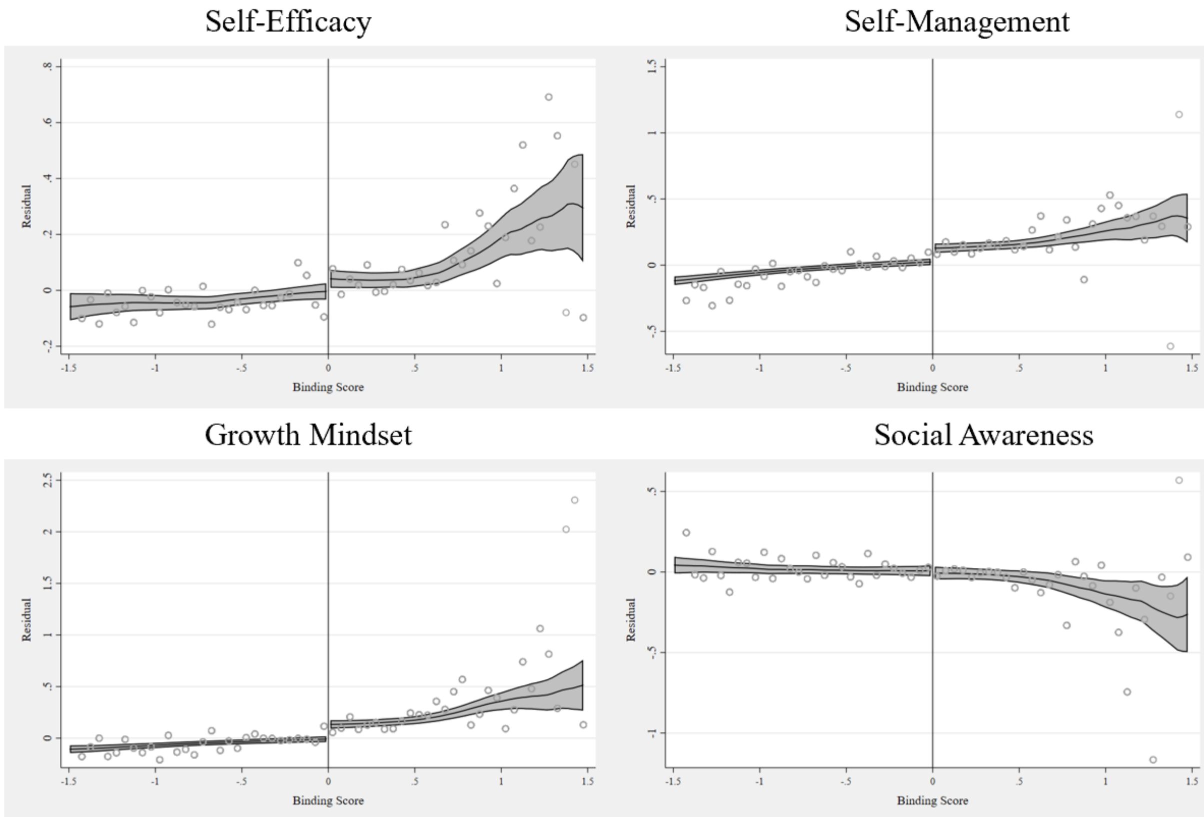
Notes: McCrary density tests run separately on 1) standardized CST-ELA centered at the cut point, and 2) binding score of the CELDT (i.e., standardized minimum value of the overall CELDT score and four subsections of the CELDT – reading, writing, speaking, and listening – centered at the cut point).

Figure S4. Covariate Balance by Binned Binding Score.



Notes: The rating variable is a student's binding score (standardized minimum value of student's CST-ELA and CELDT scores), centered at the cut point. The vertical axis is the average proportion of each respective pre-treatment demographic characteristic in each bin of binding scores. Bin width = 0.05 SD.

Figure S5: Bin Plots of SEL Residuals by Binding Score.



Notes: Each point pertains to the average value of the residual for each SEL construct for students with binding scores falling within each bin. Bin width = 0.05 SD. Gray areas are 95% CI.

Table S1. Replication of Main Model Using Optimal Bandwidth Strategy Derived from Calonico, Cattaneo, Farrell, & Titiunik (2017).

	Linear				Quadratic			
	(1)		(2)		(3)		(4)	
Self-Efficacy	0.264	**	0.294	**	0.296	**	0.302	**
	(0.090)		(0.102)		(0.107)		(0.106)	
N	4601		3770		6703		6791	
Self-Management	0.083		0.051		0.038		0.052	
	(0.111)		(0.137)		(0.141)		(0.150)	
N	4065		3005		5627		5330	
Growth Mindset	0.176	+	0.203	*	0.213		0.234	+
	(0.097)		(0.100)		(0.143)		(0.135)	
N	4613		4386		5061		5507	
Social Awareness	0.11		0.106		0.122		0.126	
	(0.111)		(0.134)		(0.125)		(0.140)	
N	4856		3715		7549		6673	
District FE	No		Yes		No		Yes	

Note: Robust standard errors in parentheses. Each coefficient represents a separate regression using the *rdrobust* command on each respective SEL outcome. Other coefficients omitted from display. All coefficients report impact at the optimal bandwidth, which ranges between .28 and .71 for the four outcomes.
+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Table S2. Robustness Check: Replication of Main Results Using Fuzzy Frontier RD Approach

	(1)		(2)	
	CST-ELA		CELDT Binding Score	
	(1)		(2)	
Self-Efficacy	0.147	***	0.18	***
	(0.044)		(0.032)	
N	8998		6686	
F	3674		10429	
Self-Management	0.141	*	0.285	***
	(0.056)		(0.037)	
N	7072		6956	
F	3333		10706	
Growth Mindset	0.211	***	0.261	***
	(0.053)		(0.034)	
N	6473		7685	
F	3224		11574	
Social Awareness	0.091	+	0.236	***
	(0.054)		(0.035)	
N	10611		10670	
F	4005		13216	

Note: Robust standard errors in parentheses. All coefficients report impact at the optimal bandwidth. Each regression coefficient represents a separate 2SLS regression. Other coefficients omitted from display. Models include district fixed effects (models without fixed effects are similar in results and are omitted from display). Column 1 shows coefficients from separate regressions using a singular rating variable, comprised of standardized CST-ELA centered at the cut point, to determine the discontinuity. Column 2 does the same using a rating variable comprised of the CELDT binding score (i.e., standardized minimum value of the overall CELDT score and four subsections of the CELDT – reading, writing, speaking, and listening – centered at the cut point).

+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Table S3. 2SLS Estimations Using Pre-Treatment Covariates as Outcome Variables.

	(1)
Female	0.041
	(0.033)
N	6963
F	2620
Latinx	0.024
	(0.018)
N	7194
F	2744
Socioeconomically Disadvantaged	0.037
	(0.040)
N	4397
F	1347

Note: Robust standard errors in parentheses. All coefficients report impact at the optimal bandwidth. Each regression coefficient represents a separate, auxiliary 2SLS regression using a pre-treatment covariate as the outcome variable. Other coefficients omitted from display. Models include district fixed effects (models without fixed effects are similar in results and are omitted from display).
+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001.

Table S4. Falsification Check Using Placebo Cut Points.

	(1)		(2)
	1 SD below 0		1 SD above 0
Self-Efficacy	-0.078		0.317
	(0.182)		(0.206)
N	6270		7953
F	395		167
Self-Management	0.023		-0.003
	(0.114)		(0.152)
N	8929		10205
F	1458		344
Growth Mindset	0.253	*	-0.324
	(0.107)		(0.229)
N	8878		8023
F	1465		172
Social Awareness	-0.064		0.321
	(0.166)		(0.288)
N	7634		7637
F	820		151

Note: Robust standard errors in parentheses. All coefficients include the optimal bandwidth. Each regression coefficient represents a separate 2SLS regression. Other coefficients omitted from display. Models include district fixed effects.

+ p<0.10 * p<0.05 ** p<0.01 *** p<0.001.