Modeling Associations of English Proficiency and Working Memory with Mathematics Growth:

Implications for RTI

By

Garret James Hall

A dissertation submitted in partial fulfillment of

the requirements of the degree of

Doctor of Philosophy

(Educational Psychology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2020

Date of final oral examination: 05/06/2019

This dissertation is approved by the following members of the Final Oral Committee:
Craig Alan Albers, Associate Professor, Educational Psychology
Stanley A. Garbacz, Associate Professor, Educational Psychology
David Kaplan, Professor, Educational Psychology
Martha Alibali, Professor, Psychology

Acknowledgements

This dissertation would not have been possible without the support of my faculty mentors, colleagues, friends, and family. I am particularly grateful for the support of my advisor, Dr. Craig Albers, and my committee. Your instruction in class, professional mentorship, constructive criticism, and personal support has been priceless. Thank you for the concerted time and energy you have all put forth to help me build my skills as a researcher and practitioner. Thank you, Carrie, for your companionship in life. Your loving support of my goals and well-being is invaluable. To my family: thank you for your endless support of me in pursuit of my education. Finally, thank you to my friends and classmates at UW for your friendship throughout the last four years. You have provided so much more meaning and value to graduate school than I could have imagined.

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Award #R305B150003 to the University of Wisconsin-Madison. The opinions expressed are those of the author and do not represent views of the U.S. Department of Education.

Table of Contents

Abstract

Although mathematics development research has become more prominent in the school psychology literature in recent years, a large research gap in this area exists regarding English language learners (ELLs). Significant research in education, cognitive science, and psychology has been devoted to understanding the links between language and mathematics, but the developmental factors that predict stability and change in mathematics skill development among ELLs are less clear. In one study, I investigated growth and change among ELLs compared to English-proficient student (EPS) peers using latent change score modeling to detect differences in interindividual and intraindividual change across language proficiency groups. I also examined the extent to which working memory differentially predicts mathematics change trends and patterns across language groups. In a second study, I closer examined the processes of mathematics development among ELLs by investigating the presence of heterogeneous, unobserved growth trajectories; whether development of English language proficiency (ELP) predicts mathematics growth patterns; and the interaction between working memory and language proficiency gains in predicting later mathematics growth. Results from Study 1 suggested that ELLs and their English-proficient peers change similarly across time though at different levels of performance. Additionally, kindergarten working memory operated similarly in predicting growth parameters among both ELL and EPS subpopulations. Findings from Study 2 are suggestive of a single highly variable growth trajectory among ELLs. Although working memory uniquely predicted mathematics developmental patterns, interactions between working memory and English proficiency gains were not predictive of later mathematics level or change pattern. However, gains in English early reading skills (arguably one domain of ELP) uniquely and positively predicted later mathematics performance level, although that advantage was partially offset by a prediction of decelerating mathematics growth through fourth grade. Theoretical and practical applications for research involving ELLs in the context of RTI, future directions, and limitations are discussed.

Keywords: *English language learners, mathematics, English language proficiency, Response-to-Intervention, working memory*

List of Tables

List of Figures

CHAPTER I: INTRODUCTION

Background

Statement of the Problem

Adequately developing mathematics skills early in schooling is important to preventing academic risk in later grades given the iterative, sequential, and interdisciplinary nature of mathematics development. Many studies underscore the importance of early mathematics competencies in predicting later mathematics skill outcomes and the developmental course of acquiring mathematics knowledge (Duncan et al., 2007; Geary, Nicholas, Li, & Sun, 2017; Jordan, Kaplan, Locuniak, & Ramineni, 2007; Jordan, Kaplan, OIáh, & Locuniak, 2006; LeFevre et al, 2010; Nguyen et al., 2016; Watts et al., 2015; Watts et al., 2017). Despite the advances in research in developmental processes in this area, less attention has been devoted to these trends among English language learners (ELLs). Recent studies have detailed how language plays a key role in much of learning mathematics and applying mathematical knowledge to problem solving across educational levels (Chow & Ekholm, 2019; LeFevre et al., 2010; Powell, Cirino, & Malone, 2017; Purpura & Logan, 2015; Purpura & Reid, 2016; Vukovic & Lesaux, 2013). Thus, it is imperative that researchers better equip practitioners with the appropriate knowledge regarding the role of language in mathematics, as effectively leveraging language use in early mathematics shows promise to benefit all students (Doabler, Clarke, & Nelson, 2016; Jayanthi, Gersten, & Baker, 2008). Given the pressing practical and policy concerns regarding the educational success and attainment of linguistically diverse students (Kieffer & Thompson, 2018; Murphy, 2014; Umansky, 2016, 2018) and paucity of research in this area, such effective instructional practices are critical to promoting students' access to high-quality instruction in mathematics (Robinsin-Cimpian, Thompson, & Umansky., 2016). Empirical research drawing on education, developmental, and cognitive science can help inform building an evidence-base in the role of language in mathematics. However, doing so is also a multidimensional process that accounts for key ecological factors such as cultural, instructional context, and policy-level factors.

Although this issue of the role of language in mathematics requires researchers and practitioners to identify levers of change at all ecological levels, a focus on student-level development of mathematics skills and language proficiency offers a bottom-up perspective to how those levers may operate at different points in students' education and brings attention to the potential developmental cascades (Lerner, Leonard, Fay, & Isaac, 2011) that may drive individual development. Such cascades are rooted in developmental systems theory (DST; Ford & Lerner, 1992; Griffiths & Tabery, 2013; Lerner et al., 2011) in which individual characteristics and contexts are theorized to interact to produce patterns of individual development (i.e., dynamic interactionism [Ford & Lerner, 1992; Lerner, 1978]). Though the current study does not directly test developmental cascades as *both* theoretical and analytic models (e.g., Berry & Willoughby, 2017; Masten & Cicchetti, 2010), there are complex differential individual-by-context interactions that take place across English-proficient students (EPS) and ELLs given differences in social context of education (e.g., placement in language instructional programs), cognitive development (e.g., learning a second language), as well as how instruction is accessed and its effect on the individual (e.g., how language instructional programs help students respond to classroom instruction). These interactions, though again not analytically tested here, are important considerations in the larger theoretical context of studying development in a dynamic ecological system such as education.

Ecological (Bronfenbrenner, 1979; Bronfenbrenner & Morris, 2006) and developmental theory helps frame how cascades may propagate multilevel and multidimensional process among ELLs and EPSs. From an empirical perspective, there is a significant need to better understand the quantitative patterns of development in mathematics across these two student groups, the factors that predict how intraindividual change transpires, and, in particular, the role of language within a mathematics developmental system. Research targeting these areas can help researchers better understand the course of mathematics development, how ecology relates to development, and what instructional and intervention practices can most effectively and efficiently produce desirable student outcomes. Recent empirical evidence provides insight into these relationships and a basis for investigating such areas of development. Purpura and colleagues (i.e., Purpura & Ganley, 2014; Purpura & Logan, 2015; Purpura & Reid, 2016) have provided evidence for language-dependent processes of early mathematics skill development among preschoolers (typically aged between four and six across studies). Similarly, LeFevre et al. (2010) proposed a model of mathematical competence in seven different areas (numeration, calculation, geometry, measurement, number line slope, number line r^2, magnitude comparison), finding that language ability measured through receptive vocabulary was the only variable that uniquely predicted each of the seven mathematics outcomes. Given that this connection of language to mathematics tends to occur regardless of native language, the language-mathematics relationship lends further credit to the multiple language demands in ELLs' learning (Baker et al., 2014; Doabler, Nelson, & Clarke, 2016): learning both general and content-specific language in academic domains. Last, these students must navigate social language development, which also bears on academic, social, emotional, and behavioral functioning (Albers, Mission, & Bice-Urbach, 2013). These language demands have been the subject of recent large-scale experimental investigations of early mathematics curricula (Doabler, Clarke, Kosty, et al., 2016; Doabler et al., 2019), mainly through probing moderators of mathematics intervention. However, the research on potential cognitive mechanisms that drive these strategies remains meager. Students who exhibit inadequate skill developmental in mathematics, coupled with acquiring a second language, may have unique factors contributing to their academic difficulties and successes. As such, it is important not only to understand general trends in the relationship between language and mathematics but also how mathematics development and second language acquisition interact over the early and intermediate elementary years.

That said, many pictures of ELL achievement trends may be painted depending on the data source. Prior work has set the precedent for highlighting the significant gaps between language groups while devoting far less attention to the patterns of those gaps over time and how individuals change within differential trajectories. However, using recent National Assessment of Educational Progress (NAEP) data, Kieffer and Thompson (2018) provided evidence that multilingual students (not limited to current ELLs) have been improving academically at a faster rate in mathematics than their monolingual peers, narrowing the achievement gap over the last decade. Despite these promising findings, significant questions about ELLs' academic development remain that require further investigation at all levels of the educational system – in the classroom, in individual interventions (i.e., single-case designs), in school districts, and at the national level. Projections of growth in science, technology, engineering, and mathematics (STEM) jobs further underscore the imperative to continue investigating mathematics among ELLs: Vilorio (2014) projected significant STEM jobs growth by 2022—some by nearly 40%. With this growth in the workforce, it is important to establish access to rigorous STEM content by leveraging the strengths of educational programming for ELLs (Umansky, 2016, 2018) and culturally relevant curricula (Driver & Powell, 2017; Dee & Penner, 2017). Increasing the positive interactions students might have with STEM disciplines can contribute to sustained interest over time (Ainley & Ainley, 2015), though it is necessary to first establish access to all instructional content among ELLs (Robinson-Cimpian et al. , 2016) to effectively garner students' learning strengths.

The Role of Response to Intervention (RTI)

The response to intervention (RTI) process can enhance practitioners' ability to provide such learning supports to students at all levels of performance. Typical models of RTI service delivery specify a strong support system at the primary (tier 1), secondary (tier 2), and tertiary (tier 3) levels (Fuchs & Fuchs, 2007). Universal screening systems and the measures employed in these systems have established a standard for the notion of risk and how schools should identify those students in need of secondary and tertiary supports (Compton, Fuchs, Fuchs, & Bryant, 2006; Glover & Albers, 2007), provided that appropriate data corroboration is conducted among school-based problem-solving teams (Albers et al., 2013). For example, Baker et al. (2019) directly studied the role of school-based data teams in identifying students in need of a school-wide, tier 2 reading intervention. In particular, they studied the process by which schools employ student achievement data to identify instructional supports. Baker et al. take note of Institute of Education Sciences' practice guide for schools in leveraging student data to identify students in need of supplemental supports (Hamilton et al., 2009). These teams and school practices, though an indirect result of the study of development, employ information on the normative patterns of student academic growth to determine intervention need, eligibility, and evidence-based preventative practices.

Yet the understanding of academic risk and the necessary supports for mathematics skill development among ELLs is less well-established, potentially making it more difficult for school-based teams to employ screening and progress-monitoring data to make highly informed decisions. First, a number of studies have elucidated the complications to assessing ELLs in their mathematics skills and the achievement gaps that seem to consistently appear between ELLs and their EPS peers (Abedi, 2002; Abedi & Herman, 2010; Abedi & Lord, 2001). Screening for mathematics difficulties among ELLs is difficult as many students already score below average benchmarks of academic performance, and coupled with small sample sizes, establishing predictive validity and norms among ELLs can be difficult (Hall, Markham, McMackin, Moore, & Albers, 2019). Albers and Martinez (2015) underscore these complications, recommending that practices such as utilizing local norms and cut scores can help reduce false rates in screening. These complications make the definition of risk in mathematics among EPSs difficult to extrapolate to ELLs. Many cultural factors may relate to these gaps, not the least of which is the language difference and the variability among those language differences.

Beyond focusing only on the language gaps, bioecological factors (Bronfenbrenner & Morris, 2006) contribute to such learning and access gaps. Often these learning and access gaps have been framed as deficits in ELLs' learning. Although many subgroups that comprise ELLs in American schools are traditionally more socioeconomically disadvantaged (Reardon & Galindo, 2009), this does not inherently imply any deficit to learning. Rather, these learning gaps arise as an interaction between the individual and the context of their development (Lerner et al., 2011). For example, Roberts and Bryant (2011) found that the school composition of students eligible for free and reduced lunch explained a significant amount of variance in mathematics development across English-speaking students as well as Spanish and Asian ELLs. Abedi and Herman (2010) found that the proportion of other ELLs in the classroom explained significant variance in mathematics outcomes. Such contextual factors interact with individual students' response to instruction, including access to and effective learning of grade-level content and progress comparable to peers according to both national and local norms.

These data, both quantitative and qualitative, provide the basis for instructional supports, intervention identification, and, absent of student responsiveness to tiered intervention, special education eligibility. However, many empirical studies further obscure potential variation in patterns of ELL mathematics development by classifying "ELL" as a binary indicator when there is substantial variability along the language proficiency continuum. Additionally, how that binary indicator is defined varies significantly in research. Although this binary formulation of ELLs is efficient and is in some cases the only information available, it is clear this approach likely suppresses substantial variance within ELLs that could more effectively elucidate developmental and achievement trends among this group.

In translation to practice, these varying definitions have significant implications for access to instruction required to perform at grade-level (Robinson-Cimpian et al., 2016) and the cognitive factors implicated in second language learning (Bialystok, 2011; Greenfader, 2017; Grundy & Timmer, 2017; Wang, 2017). For example, these variations in English proficiency, especially the timing of obtaining proficiency, may have different implications at different grade levels (Halle et al., 2012). In research, definitions of ELLs and the subsequent decomposition of between- and within-group relationships have epistemological implications of research among ELLs: How do we actually know what we think we know about ELL achievement, and what is the relationship of researchers' methodology to generating that collective understanding? As such, research must address the individual differences within language minority groups with greater attention to the characteristics that define such groups as well as to development of English language proficiency (ELP). This is especially important in mathematics in light of the significant research gaps in mathematics development among ELLs.

Patterns of Development and RTI

The topography and distribution of developmental trends is particularly emphasized here due to their role in forming a comprehensive, preventative RTI model. Studying variability in development can help extract students who differentially change in systematic ways. Such sub-distributions might include "movers" and "stayers" (Goodman, 1961; Kaplan, 2008). Varying definitions could be formulated for the mover-stayer framework in education. One such definition might include students who respond (i.e., increasing level and trend) or do not respond to instruction (i.e., low level and minimal trend), a pattern that Jordan and colleagues (2006; 2007) partly observed in early numeracy skills. This may appear qualitative based on observing the predicted trajectories, or the longitudinal likelihood of switching trends or categories can be empirically estimated (Chow, Dolan, Grimm, & McArdle, 2013; Kaplan, 2008; Kaplan & Walpole, 2005). Extracting these subgroups of growth and change may help derive levers of intervention and assessment in RTI. Due to the unique needs of ELLs, however, a more thorough picture of mathematics development among ELLs must be painted by examining shapes, patterns, and distributions of mathematics development.

The Present Work

The two studies included herein address gaps in the literature on ELL development in mathematics with particular emphasis on how studying developmental topography can inform RTI practices and future research on assessment and intervention in mathematics among ELLs (as well as EPSs). Across the two studies, I address development with a focus on patterns of variability and shape. Many prior studies have examined the persistent achievement gaps among ELLs; however, given the heterogeneity among ELLs, fewer studies address differential patterns of change between ELLs and EPSs, within the ELL subgroup, and the role of working memory and language proficiency in predicting these relationships. As RTI becomes more crucial to school-based policy, access to grade-level content must be addressed from multiple angles, rather than focusing on average gaps between demographic groups. Among ELLs in particular, this means analyzing academic skill and curricular access along the ELP continuum. Recent work has exemplified the efficacy of tier-2 mathematics intervention among ELLs with potentially differential benefits for ELLs with lower ELP (Doabler et al., 2019). Additionally, working memory is a stronger predictor of later mathematics achievement among Hispanic students who exhibit fewer basic reading skills in English (Greenfader, 2017). Though such studies are very informative for the role of cognitive skills among ELLs and responsiveness to mathematics instruction, these studies do not directly address how ELP relates to differences in intraindividual change compared to EPSs or later developmental patterns in terms of the distribution of trends and variability in patterns of change. Furthermore, few studies have addressed the relationship between ELP and mathematics as sequential development processes. Halle et al. (2012) provided cogent evidence for the importance of the timing of obtaining ELP, and as such, the rate at which early ELP develops may play an important role in predicting the subsequent pattern of mathematics growth, especially when students are assessed in English. The present work addresses this research gap.

Breakdown of Studies in Current Research

Study 1. In Study 1, I extend prior work (Roberts & Bryant, 2011) analyzing the developmental trend of mathematics between ELLs and EPSs. I take a latent change score modeling approach (McArdle & Nesselroade, 2014) to analyzing differential patterns of growth between low-ELP ELLs and English-proficient students while controlling for a variety of behavioral, cognitive, demographic, and ecological factors. I investigate the invariance of change patterns across groups to detect whether low-ELP ELLs and EPSs change in significantly different ways, both in the constant amount of change that accrues between kindergarten and fourth grade as well as the proportional change that occurs between grade levels. Additionally, given the role working memory plays in mathematics development, I test whether working memory differentially predicts mathematics development patterns across groups. In Study 1, I address the following research questions:

Research Question 1: Does the pattern of mathematics growth from the Spring of kindergarten to the Spring of grade 4 differ significantly for ELLs and EPSs in terms of additive and proportional change?

Research Question 2: To what extent does working memory differ across ELLs and EPSs in uniquely predicting mathematics development patterns?

Study 2. In Study 2, I focus on the students I defined as low-ELP ELLs in Study 1 and explore the presence of heterogeneous patterns of growth in mathematics using growth mixture modeling (Kaplan, 2002; Ram & Grimm, 2009). Contingent upon the presence of mixture patterns, I also planned to test whether students are stable in their growth trajectory membership across time (i.e., regime switching; Chow et al., 2013). Again, contingent on the presence of mixtures, I also planned to test whether early growth in ELP predicts growth trajectory membership and the stability in growth class membership. To further investigate the relationship between working memory and ELP, I also planned to test whether working memory moderates the extent to which ELP growth predicts later mathematics outcomes, including growth class membership, growth class stability, and the growth patterns within each mixture class. If mixture classes exist, testing these relationships can help identify heterogeneous relationships within and between unobserved growth patterns. Recent research has advanced in testing differential response to intervention based on ELP (Doabler et al., 2019), working memory ability (Fuchs, Schumacher, Sterba, et al., 2014), and baseline mathematics proficiency (Fuchs, Fuchs, & Gilbert, 2018). As such, it is necessary to detect when and for whom ELP gains predict growth patterns to better inform relationships between language and mathematics development and instructional response. The research questions for Study 2 are as follows:

Research Question 1: Are there unobserved, systematic, differential growth trajectories (i.e., mixtures) in mathematics development from grades 1 to 4?

Research Question 2: Do gains in ELP across kindergarten predict greater mathematics growth? If multiple growth mixtures exist, do ELP gains predict growth trends differently within each class?

Research Question 3: Is ELP is a stronger predictor of mathematics growth patterns among students with lower working memory capacity?

CHAPTER II: LITERATURE REVIEW

RTI is predicated on an understanding of developmental cascades (Ford & Lerner, 1992; Lerner et al., 2011) and the ecology of students' development (Burns, 2011). RTI helps schools identify students at varying levels of multidimensional risk as early as possible and provide services to prevent further academic performance challenges. Such a process requires assessments to measure the developmentally appropriate skills and interventions to target the appropriate cognitive mechanisms of learning. Most importantly in the RTI model, those assessments and interventions contribute to preventing a particular issue from compounding over time, which means that it is critical to understand the developmental topography of that skill domain. Specifically, translational researchers must understand the deficits and challenges that may accrue and cascade in absence of intervention (i.e., counterfactuals); how intervening on a particular skill aligns with instruction in subsequent curriculum units and grades; and the risk and protective factors that contribute to those developmental cascades. The unique strengths of ELLs and challenges to their success necessitates a stronger conception of how RTI can be most effectively leveraged for their development. Doing so requires a developmental approach to studying mathematics skills among ELLs, specifically focusing on the differential patterns that may emerge in light of the distinctive linguistic and cognitive factors related to ELLs' mathematics learning.

Characteristics of English Language Learners and Their Education

Demographic and Geographic Characteristics

ELLs represent a large and growing population of students in the United States, comprising 9.6% of students in elementary and secondary schools in 2016 (U.S. Department of Education, National Center for Education Statistics, 2018a). In early elementary grades, ELLs comprise up to 16% of the total student population (U.S. Department of Education, National Center for Education Statistics, 2018b). Bardack (2010) defined an ELL as "…an individual who is in the process of actively acquiring English, and whose primary language is one other than English" (p. 5). Although Spanish-speaking students represented the majority of ELL students in American schools in the 2016 school year, totaling 77% of ELL students (U.S. Department of Education, National Center for Education Statistics, 2018b), the general population of ELLs is in fact extremely diverse, representing many different native languages from across the world. Other commonly spoken languages include Arabic, Chinese, and Vietnamese, with each of those languages representing around two percent of those spoken in American schools (U.S. Department of Education, National Center for Education Statistics, 2018b). The regions these languages represent range from the Middle East, to Southeast Asia, Sub-Saharan Africa, Central Europe, and to regions within the United States, where some students speak primarily Native American languages (U.S. Department of Education, National Center for Education Statistics, 2018b). Moreover, the countries and regions from which these students (first generation) or these students' parents and families (second generation) originate vary significantly in religious and political affiliation as well as economic and educational environment; Reardon and Galindo (2009) highlight this diversity among Hispanic students in particular. As a result, there is significant heterogeneity within the ELL population. Some subgroups of the ELL label may be more socioeconomically marginalized than others (as Reardon and Galindo [2009] highlight among Hispanic students), making meeting the needs of these students challenging, particularly in areas with relatively few resources for culturally diverse or immigrant families (e.g., bilingual specialists in schools). Thus, maximizing these students' strengths and attainment rests not only on language but also on competence and understanding of the cultural factors that impact their learning (Albers et al., 2013).

This demographic of students has been growing steadily over the past ten years, especially in areas of the American West and South (U.S. Department of Education, National Center for Education Statistics, 2018a). Additionally, the number of students participating in programs designed to serve ELL students varies significantly by state and school district. For example, California has the largest population of ELLs in the United States, totaling 1,261,672 students across K-12 education in 2016 (U.S. Department of Education, National Center for Education Statistics, 2018a). This number has fallen by nearly 300,000, however, since 2005, when 1,571,463 students were enrolled in ELL programs (U.S. Department of Education, National Center for Education Statistics, 2018a). Conversely, since 2000 ELL enrollment has been steadily climbing in other states. In Texas, ELL program enrollment totals have increased by more than 100,000 students since 2000; in Florida, by 80,000 students. From 2000 to 2016, North Carolina's ELL program enrollment total more than doubled from around 44,000 to around 92,388 students. In Arkansas, enrollment more than tripled from around 12,000 to 41,482 students between 2000 and 2015. In many states, this growth is expected to continue through 2027. For example, from Fall 2015 to Fall 2027, ELL enrollment rates are expected to increase by up to 27% in some states (e.g., North Dakota), while in many others states ELLs are expected to increase enrollment between 10% and 17% by 2027 (McFarland et al., 2018). Enrollment numbers are expected to decrease in other areas, including states with already high numbers of ELLs, such as California and Illinois (McFarland et al., 2018).

Educational Resources, Instruction, and Programming

The prior statistics are important not only to show the major population growth and fluctuations; they are also important indicators of the educational resources needed to adequately serve these students. Many areas of extreme growth may not have the material (e.g., curricular materials, classroom space) or human resources (e.g., administrators, social workers, school psychologists, bilingual educators) to keep up with the immense growth. Such contextual factors may explain differential opportunities to learn for ELLs. For example, Abedi and Herman (2010) found the classroom-level opportunities to learn in eighth grade algebra courses were higher within classrooms with a higher proportion of ELLs. Furthermore, the resources that are in place may not be necessarily culturally aligned to garner these students' academic and social-emotional-behavioral success (Robinson-Cimpian et al., 2016; Umansky, 2016). These ecological factors highlight not only the need to understand the context itself but also how development occurs in context, inevitably shedding light on the crucial individual-by-context interactions that drive developmental cascades (Bronfenbrenner & Morris, 2006; Lerner et al., 2011). With this in mind, attaining ELP can be a challenging process both for the student and their teachers. Many classrooms might have students with varying levels of English proficiency in addition to a wide variety of academic skill level and social-emotional-behavioral difficulties, all of which may impact access to learning opportunities among ELLs. Unlike a completely English-speaking classroom with a variety of achievement levels, adding a language barrier could significantly affect a teacher's ability to target skill areas in which an ELL may be struggling (e.g., how a student might conceptualize a problem differently based on language; cultural differences affecting what knowledge teachers assume students have). That said, value-added modeling approaches (Loeb, Soland, & Fox, 2014) have shown that teachers who are generally effective with EPSs generalize their skills to working with ELLs. Although, Loeb et al. (2014) and Master et al. (2012) both found that additional training for working with students provides an advantage in producing positive academic outcomes among ELLs.

Reaching language proficiency is a complex and multidimensional process, which may explain why additional bilingual training among teachers may be so crucial for student outcomes. Language needs to be attended to at both the social and academic level (Albers et al., 2013). Proficiency in social language does not constitute proficiency in academic language, as academic language is content- and task-specific. As such, teachers and instruction more generally must be aligned with the developmental context, that language skills among ELLs are tied to more than just the content-specific language required to learn the target skill. Simultaneously, language is intuitively embedded in many mathematics abilities, with word problem solving being the most clearly language-embedded activity. That said, language underlies mathematics skills very early on (Purpura & Reid, 2016). Purpura and Reid (2016) suggested that poor understanding of mathematics vocabulary in preschool and the relationships among mathematical concepts portends poor performance in kindergarten, as these linguistically embedded concepts of numeracy underlie the procedures and concepts needed to build early number sense. Nevertheless, both social and academic language are culturally embedded, which is not uniform across student backgrounds. Development of proficiency in these English language domains is supported in schools through a number of different types of programs, which are discussed below.

ELL education programming. A variety of language instruction programs are employed throughout the United States to support primary (native) languages (L1) and second language (L2) proficiency. Faulkner-Bond et al. (2012) described seven predominant language instruction education programs (LIEPs). The instructional program orientation and the goals of each program are listed in Table 1 below.

Table 1
ELL instructional programs and language goals from Faulkner-Bond et al. (2012).

Program Type	Program Goal
English as a second language (ESL) instruction	"Proficiency in English" (p. x)
Content-based instruction	"Preparation to meet achievement standards, proficiency in English" (p. x)
Sheltered instruction	"Preparation to meet academic achievement standards, proficiency in English" (p. xi)
Transitional bilingual programs	"Preparation to meet academic achievement standards, proficiency in English" (p. xi)
Two-way immersion	"Preparation to meet achievement standards, proficiency in English, bilingualism and biliteracy, biculturalism" (p. xii)
Developmental bilingual programs	"Preparation to meet academic achievement standards, proficiency in English, bilingualism and biliteracy" (p. xii)
Newcomer	"Preparation to participate in regular LIEP offerings, build foundational skills in content areas (basic literacy, math concepts, etc.)" (p. xiii)

The programs that are employed in schools vary by state and district policies. In most states, there are a variety of programs offered depending on district resources. Rolstad, Mahoney, and Glass (2005) conducted a meta-analysis that examined ELL education programs and revealed that bilingual education was more effective in producing higher achievement when comparing ELLs to other ELLs ($ES = 0.23$). Umansky (2016) utilized a regression discontinuity design (RDD) to evaluate the impact of language instructional programs in kindergarten through high school, showing similar results as Rolstad et al. (2005). In particular, Umansky (2016) showed that bilingual (as well as dual-language immersion) education programs were most impactful in producing achievement in mathematics and language arts for ELLs who fell just below the cutoff for language support eligibility. However, the implications of such an effect become highly muddied as one infers further away from the discontinuity cutoff, underscoring the need to better understand how achievement and policy initiatives interaction among students in a wide range of ELP levels.

Recent policy changes have significantly altered the ELL education landscape by aligning with recent empirical evidence. In November 2016, Proposition 58 in California (Senate Bill 1174) was approved in their state election, which changed their ELL instruction policy to include non-English languages, having previously instructed all students in English (California Education for a Global Economy Initiative of 2014). This reflects a cultural shift among policymakers that recognizes, although English language proficiency is a crucial mechanism to academic success, the process of building that process recruits many processes of development, including proficiency in native language (Albers & Martinez, 2015; Cárdenas-Hagan, Carlson, & Polard-Durodola, 2007; Cummins, 1979; Guglielmi, 2012). Additionally, different second language programs may have different social psychological implications for students and may be differentially resourced in terms of providing for students (Umansky, 2016). These contextual factors are undoubtedly important to consider in light of the goal of language support services: helping students master grade-level knowledge domains. However, such contextual factors that filter into the psychological processes of learning may relate to students' learning. This is not the focus of the current work, although the overarching ecology of ELLs' education experiences underscores much of the current and extant work in this area.

For language instructional programs to work successfully, it is necessary for practitioners to build students' proficiencies in a way that allows them to access grade-level instruction to meet these standards. However, the supports within different types of language programming may have significantly different mechanisms of support students' instructional access and instructional response that are related to the cognition of instruction, the teachers' training, and the social context of the instructional environment. As Umansky (2016) highlighted, placement in different kinds of language instructional classrooms may have significantly different implications for students' social experiences of their education. These types of instructional contexts may act on different mechanisms of the relationship between L1, L2, and mathematical cognition. At a more basic level, teachers and interventionists must attend to a student's language proficiency when accounting for their response to instruction, and the extent to which language proficiency facilities or impedes such response. With policy-level standards in mind, it is necessary to understand how mathematical knowledge develops, its relationship to language, and the multilevel factors that promote or impede students' capacity to meet grade-level benchmark proficiencies.

Properties of Mathematics Development

Many studies empirically describe language-mathematics associations in school-aged children learning second languages (Abedi & Lord, 2001; Chen & Chalhoub-Deville, 2016; Garrett, 2010; Greenfader, 2017; Guglielmi, 2012; Halle et al., 2012; Hartano, Yang, & Yang, 2018; Roberts & Bryant, 2011; Van Rinsveld, Schiltz, Brunner, Landerl, & Ugen, 2016; Van Rinsveld, Dricot, Guillame, Rossion, & Schiltz, 2017; Vukovik & Lesaux, 2013; Wang, 2017). Chen and Chaloub-Deville (2016) found that much of the achievement discrepancy between ELLs and EPSs disappears when language proficiency (as measured by English reading ability) is controlled; however, they also noted that there is a large amount of variability within language groups that seems to be driven by language demands. Additionally, the same authors found the relationship between academic language proficiency was stronger across time for students with lower mathematics achievement than for students performing higher in mathematics . Thus, although language ostensibly impacts ELLs most significantly—and these students are in the most need for language support—the language demand of mathematics impacts all students. Other studies in RTI, as well as broader developmental theory, corroborate the role of general language ability. From a tiered intervention perspective, Powell et al. (2017) showed that language ability uniquely predicted responsiveness to a tiered mathematics intervention for students at-risk in mathematics. LeFevre et al. (2010) found similar evidence of linguistic ability uniquely predicting mathematics domains among four- to seven-year-old native English-speaking students. Among ELLs, Halle et al. (2012) found that earlier attainment of ELP in kindergarten was predictive of a steeper growth trajectory in mathematics compared to students who reached proficiency in first grade. Both Greenfader (2017) and Wang (2017) used the ECLS-K: 2011 dataset to investigate mechanisms of achievement patterns among Hispanic students. The results of these studies suggested not only enduring relationships between English proficiency and mathematics achievement, but unique pathways to mathematics knowledge that are consistently related to language across developmental and competency levels in mathematics and native language.

Mathematics Cognition and Language

The mechanisms behind relationships between mathematics skills and language continue to prove somewhat elusive, although more recent research is strengthening the connection, particularly in the early childhood education context (LeFevre et al., 2010; Pupura & Ganley, 2014; Purpura & Logan, 2015; Purpura & Reid, 2016). Recent experimental work has even established a causal relationship between mathematics language and mathematics performance among preschool (Purpura, Napoli, Wehrspann, & Gold, 2017) and kindergarten children (Hassinger-Das, Jordan, & Dyson, 2015). These experimental studies have focused on training students in mathematical language; however, language has a much broader cognitive role in mathematics beyond the domain-specificity of mathematics vocabulary.

Much of the understanding of the basic number system comes from the triple-code model (Dehaene, 1992), which posits three independent "codes" of number representation. The first two codes relate to the storage of number in verbal and Arabic format. These two formats support the functioning of the exact number representation system and support exact, symbolic arithmetic functions. The third code, the approximate number system (ANS), supports representation of magnitude and quantity. The ANS develops independently of language and is a shared quality with non-human mammalian and non-mammalian mathematical cognition (Dehaene, 2011). Conversely, the verbal and Arabic codes are supported largely through language and the mapping of approximate quantities to discrete, exact representations of number.

Given the emphasis on the phonologically- and symbolically-encoded exact number representation codes, language is potentially a high-leverage factor in the development of exact arithmetic skills. These skills are the basis of mathematical development in formal school (though the early ANS system alone is predictive of school-aged mathematics ability [Mazzocco, Feigenson, & Halberda, 2011] and the interrelationship of exact and approximate number representations potentially lies at the center of specific learning disorder in mathematics [Noël & Rousselle, 2011]). For example, Nys, Content, and Leybaert (2013) recruited a sample of students with specific language impairment (SLI) as well as non-SLI controls to compare on approximate and exact calculation tasks. As the authors note, their findings were consistent with the triple-code model: approximate number functions remained in-tact among SLI participants, though they exhibited consistent deficits in exact number functions compared to controls. This suggests that deficits in exact number arithmetic may be at least in part attributable to language ability. Other research offers more general insight; Gelman and Butterworth (2005) noted that "… it is one thing to hold that language facilitates the use of numerical concepts and another that it provides their causal underpinning" (p. 9). This viewpoint is consistent with many of the ideas advanced in Dehaene (2011), specifically that language promotes the syntactical and functional purposes of numbers. Recent evidence confirms this role of the language structure in potentially facilitating mathematics learning (Chow & Ekholm, 2019), providing alternative evidence to the well-established role of general language vocabulary in predicting mathematics performance by showing that language syntax, over and above multiple other domains of language, is the strongest predictor of latent mathematics performance (comprised of equivalence, problem-solving, and calculation measures) among elementary students ($\beta = 0.18$, $p < .05$). Nevertheless, as much as language may mediate and/or moderate the extent to which learners are able to categorize, synthesize, calculate, and ultimately remember mathematics functions, language on its own does not hold the key to the basic number system. Yet it does play a vital role in acquiring and storing numerical information (Dehaene, 2011; LeFevre et al., 2010) and understanding mathematics (Vukovic & Lesaux, 2013). The evidence as to whether that occurs through accessing the general structure (Chow & Eckholm, 2019) or vocabulary (LeFevre et al., 2010; Vukovic & Lesaux, 2013; Vukovic et al., 2014), however, requires further scrutiny.

To investigate the specific role of language in mathematical knowledge, researchers often study bilingual-proficient individuals to analyze the extent to which language mediates and/or moderates mathematics learning and problem solving (Dehaene, Spelke, Pinel, Stanescu, & Tviskin, 1999; Marsh and Maki, 1976; Spelke & Tsivkin, 2001; Venkatraman, Siong, Chee, & Ansari, 2006). In their seminal work, Marsh and Maki (1976) recruited Spanish-English bilinguals to solve simple addition problems in either their preferred or non-preferred language. They found that the speed with which participants responded with answers was quicker in the preferred language; however, the mechanism through which this was hypothesized to occur was translation. For these participants, it was quicker to translate the arithmetic processes into the preferred language rather than answer in the non-preferred language. Spelke and Tsivkin (2001) conducted research similar to Marsh and Maki (1976) but additionally distinguished between exact and approximate number calculations. Russian-English bilingual individuals who were trained to answer exact number arithmetic (as opposed to approximate) calculated their answers quicker in their native language in comparison to their second language. When these students answered questions involving approximate arithmetic, the speed of calculation was more or less equal between the two languages. Venkatraman et al. (2006) used similar methods of exact versus approximate arithmetic calculations to study the neurological processes underlying these differences in the level of language dependency of numerical computations. This study provided evidence for a biological basis to the nature of how language relates to arithmetic. Specifically, there might be independent brain regions implicated in language switching of exact and approximate arithmetic. Van Rinsveld et al. (2016) examined arithmetic and bilingualism among secondary and post-secondary students (seventh grade through early college). Their results suggested that when students were primed with a "language context" (i.e., being asked to evaluate a true/false question) in their second language, they solved an arithmetic problem faster than without the language context ($\eta^2 = .067$). However, this effect did not hold for students' primary language: the presence or absence of a language context did not alter the speed with which the arithmetic was completed ($\eta^2 = .001$). Additionally, complex problems were completed quicker with the language context than without ($\eta^2 = .065$), though this language effect did not hold for simple problems ($\eta^2 = .001$). Across these studies, both the nature of how the problem is solved (i.e., exact or approximate solving), the language of problem solving and the conditions under which these problems are solved (e.g., the language context) appear to influence, or at the very least relate to, simple arithmetic. These factors have important instructional implications in that language has inconspicuous effects on tasks that are, on face the of it, seemingly not rooted in language.

Van Rinsveld et al. (2017) found no language differences in solving simple addition problems, although students performed better at complex addition in their native language (German) than second language (French). This line of research suggests multiple ways in which second-language learners (specifically bilingual students) encode mathematical information, and the form of long-term storage of such information seems to moderate the relationship between language of instruction/stimuli and mathematics performance. Van Rinsveld et al. specifically suggested that for simple arithmetic, the manner in which the mathematical content is encountered modulates the form of encoding. In the context of Luxembourgish schools in which students learn both German and French, the authors highlighted the likely pattern of that students learn German earlier than French (and thus learn mathematics in German), although students learn mathematics in French likely out of necessity (Van Rinsveld et al., 2017, p. 24). The type of problem-solving in which students are prompted to engage then moderates the relationship between language, type of encoding, and behavioral mathematics performance, as indicated by students' better complex addition performance in German compared to French. Specifically, the authors drew on prior fMRI work (Tang et al., 2006) to conclude that the language in which arithmetic information was encoded (e.g., in German) may subsequently affect how such information is retrieved (e.g., in French).

Distinguishing among types of mathematical cognition, Vukovic and Lesaux (2013) found that language does not play a significant role in the cognitive processes affecting strictly numerical cognition (e.g., arithmetic) among six- to nine-year-old children. However, they provided evidence that language may promote understanding of mathematics, a trend that they observed across both ELL and EPS groups. Thus, the language-based information used to solve mathematical tasks might be more heavily rooted in the underlying understanding that students generate rather than the overlap of numerical and linguistic cognition. Vukovic and Lesaux's (2013) findings, however, do not provide evidence of neural mechanisms of language differences as Van Rinsveld et al. (2017) and Venkatraman et al. (2006) do. As such, it is difficult to determine the extent to which language facilitates comprehension of mathematics, as the language of instruction and language of mathematics learning may moderate the role of language in acquiring mathematical knowledge. Language as a mechanism for comprehending mathematics may theoretically account for an equal amount of variance or similarly predict in mathematics performance across linguistic groups, though the paths through which this variance is explained may differ between groups (i.e., there are different variables mediating the theoretical link between language and mathematics performance).

Taking this evidence together, it is unclear what might be the source of this variation in cognitive storage and representation of mathematical information. The manner by which mathematics knowledge is encoded potentially varies across language groups, which then would potentially alter the manner by which students leverage their mathematical knowledge to solve problems. Based on the developmental research on the exact and approximate number systems and the role of language in developing basic mathematical representations, it would seem that the capacity to maximize exact number functions as an extension of the ANS can be in part attributable to language, regardless of whether individuals are monolingual, bilingual proficient, or learning a second language. However, given the relationships identified between language and exact versus approximate arithmetic among bilinguals, ELLs may at times encode mathematical content in a perceptually different manner. By extension, it seems intuitive that the manner by which these number representations are utilized to execute tasks in either L1 or L2 may depend on the extent to which L1 and L2 influence the storage of mathematical information. That said, because of the overlap of language and culture, there are many socioecological factors tied to language groups that undoubtedly relate to students' mathematical performance, development, and cognition.

Confounders of Mathematics and Language Development: Socioeconomic Status and Social Context

Although many studies have detailed bilingual mathematics problem solving, many of these studies (e.g., Van Rinsveld et al. 2016, 2017) conducted their research among a select group of bilinguals that did not represent the educational, social, or developmental context of second language learners (specifically, ELLs) in American schools. As such, many studies focusing on ELLs in American education account for social and economic indicators in their analysis, especially when utilizing large datasets. For example, Roberts and Bryant (2011) found that status (SES) may be a more effective predictor of mathematics achievement than language, at least throughout elementary school. These authors found that student-level SES was a better predictor of mathematics achievement than L1, with school-level SES, measured as the proportion of free-reduced lunch-eligible students attending the school (as a proxy for school resources and context), accounting for a significant proportion of mathematics achievement growth (not separated by language grouping). However, the important limitation to this work is that the authors included only students proficient in oral English at the end of kindergarten or reached proficiency during the course kindergarten. This is an important distinction in how this research fits into the broader literature, in that non-English proficient students are a main focus of educational programming and academic intervention. Although SES is a potent predictor according to research, it may not be any better of a predictor than other barriers to academic performance, such as low English proficiency, given that English proficiency is a key gateway to accessing academic content. Nevertheless, Halle et al. (2012) also showed that multilevel and multifaceted characteristics predicted the timing of reaching English proficiency (at kindergarten entry versus end of first grade), such as classroom environments (e.g., amount of time spent on reading), language barriers inhibiting home-school connections, parent education level, and teacher training in child development. These data underscore the ecological variables underlying second language proficiency development and the extent to which context interacts with individual characteristics to drive developmental patterns. Although these socioeconomic and sociocultural factors are vital to understanding the ELL demographic and their learning characteristics, they are not the sole drivers of change. To that end, it is important to study the factors that can be targeted directly within instruction and the school context (e.g., cognition, behavior) to improve mathematics outcomes considering these are the main factors driving the RTI model while recognizing the broader socioecological constructs that relate to ELLs' mathematics development.

Domain-Specific Predictors of Mathematics Achievement

A number of studies identify the early predictors of mathematics achievement, although few of them directly address language as a developmental predictor. That said, this research is useful in understanding the nature of the domain-specific (i.e., early mathematics) indicators of performance in mathematics. To start, Duncan et al. (2007) conducted a well-known study across six large longitudinal datasets indicating that mathematics, with an effect size of 0.34, is the strongest predictor of later academic achievement. Other studies examine more specific subskills of mathematics that are predictive of positive mathematics skills growth. Jordan et al. (2007) investigated the predictive value of number sense in kindergarten on first-grade mathematics achievement. Their findings indicated that growth in number sense (i.e., counting, knowledge of numbers, and simple addition/subtraction) across kindergarten and first grade, in addition to number sense skills in kindergarten, explained a large portion of mathematics achievement in first grade ($R^2 = .66$). Extending the predictive value of number sense, Locuniak and Jordan (2008) found that kindergarten number sense predicted second-grade calculation abilities more effectively than demographic and cognitive characteristics or reading abilities ($R^2 = .42$).

Comparing the development of different mathematics abilities, Fuchs, Powell, Cirino, et al. (2014) conducted a randomized controlled trial of a second-grade calculation skills intervention and a word problem solving intervention to investigate whether calculation or word problem solving abilities contributed differentially to the development of pre-algebraic knowledge. Their findings suggested that word problems significantly contributed to the development of mathematics skills compared to a control ($ES = 1.36$) and the calculation intervention ($ES = 1.32$). Interestingly, the calculation intervention did not appear to predict pre-algebraic knowledge more effectively than the control condition. Relatedly, equals sign knowledge significantly predicted two measures, solving equations and functions, of fourth grade algebra knowledge (Matthews & Fuchs, 2018), although again the extent to which general language or ELP among ELLs predicated equals sign knowledge was not directly examined. Other earlier childhood predictors of late elementary mathematics (grade 5) performance included counting and cardinality, patterning, and geometry, although counting and cardinality emerged as the strongest predictor of grade 5 mathematics performance ($\beta = 0.42$; Nguyen et al., 2016). However, Watts et al. (2017) found that a preschool curriculum designed to increase counting, geometry, and patterning skills did not directly improve stable aspects of mathematics skills over time; rather, it increased mathematics performance directly after the intervention and then faded out beyond that period. Together, these pieces of evidence suggest that different mathematics content areas are differentially longitudinally predictive. Across studies, basic number sense emerges as one of the strongest predictors of long-run mathematics achievement, with fraction knowledge and word problems representing the next steppingstone in the acquisition of core mathematics skills. Although these studies are highly informative, few of them directly address early longitudinal predictors of mathematics among ELLs.

Domain-General Predictors of Mathematics Performance and Relationships with Language

Empirical studies consistently point to working memory as a key developmental predictor of mathematics skills (Bull, Espy, & Wiebe, 2008; Geary, 2011; Geary, Hoard, Nugent, & Bailey, 2012; Cragg, Keeble, Richardson, Roome, & Gilmore, 2017; Geary et al., 2017**;** Lee & Bull, 2016) as well as a moderator of mathematics intervention (Fuchs, Schumacher, Sterba, et al., 2014). Based on initial work by Baddeley and Hitch (1974), Gathercole, Pickering, Ambridge and Wearing (2004) established invariance of a three-factor model of working memory (which includes central executive, phonological loop, and visuospatial sketchpad elements) across ages six to 15. These factors, both independently and in concert, support the short-term holding and processing of information. Working memory (both verbal and visuospatial elements) supports the development of mathematics fact fluency and problem-solving procedures, whereas verbal working memory simultaneously predicts conceptual mathematics understanding (Cragg et al., 2017). Cragg et al. additionally found that factual, procedural, and conceptual understanding also mediate the relationship between verbal working memory and general mathematics achievement over and above a direct relationship between verbal working memory and general mathematics achievement. As the authors noted, the contribution of inhibition and shifting as components of executive functioning to components of mathematics skills as well as general achievement largely disappeared when controlling for working memory, indicating working memory is a critical cognitive mechanism in supporting multiple domains of mathematics learning processes (facts, procedures, and concepts) and achievement.

However, only a handful of studies address the relationship between working memory, language proficiency, and mathematics achievement (Greenfader, 2017; Swanson, Kong, & Petcu, 2018; Wang, 2017). Building bilingualism and fostering second-language learning constitute a significant cognitive undertaking (Chen & Li, 2008; Cummins, 1979). Supporting language learning is challenging at student, instructional, and policy levels, yet some studies show that early bilingualism presents a significant cognitive advantage over monolingualism (Bialystok, 2011; Bialystok & Martin, 2004). This may suggest that there are distinct cognitive functions that relate to language learning, though the cognitive benefits of early bilingualism have come under greater scrutiny of late (Lukasik et al., 2018). Studies supporting cognitive advantages to bilingualism show that strong bilingual abilities at an early age predicts inhibition ability, likely due to the competing demands of switching between languages and inhibiting one linguistic process in favor of another (Bialystok, 2011). Other studies have focused on whether the “bilingual advantage” extends to all dimensions of executive functions. Grundy and Timmer (2017) conducted a meta-analysis to establish the consistency of findings of bilingual advantage in working memory, establishing that being bilingual presented a moderate ($ES = 0.20$) advantage in working memory ability. This study replicated a prior meta-analysis (Linck, Osthus, Koeth, & Bunting, 2014), establishing *ES* of 0.255 for the bilingual advantage in working memory ability, which was consistent across levels of bilingual proficiency, though the complexity of the working memory task (e.g., simple versus complex span tasks) partially moderated this effect. Grundy and Timmer (2017), Linck et al. (2014), and Lukasik et al. (2018) mainly focused on bilingual adults, although this potential bilingual advantage has been shown to occur early on. As evidenced by computation tasks involving multiple languages (Spelke & Tviskin, 2001), there is a nuanced relationship between mathematical tasks and language whereby language does not necessarily account for the ability to conduct mathematical operations, though it appears that much of this information needed to execute these tasks travels through language-based cognitive mechanisms (Gelman & Butterworth, 2005; Van Rinsveld et al., 2016; Van Rinsveld et al., 2017). As such, meaningful group differences between bilingual and monolingual children in executive functions may have significant implications for mathematics development, although the role of working memory in particular requires further examination.

Working memory may also be a plausible actor in the development of mathematics skills among ELLs for theoretical reasons, not just intuitive reasons based on prior empirical findings. Cognitive load theory in education research (Sweller, Ayres, & Kalyuga, 2011) suggests that instruction and learning that strains working memory capacities impedes skill development (de Jong, 2009; Sweller, 1988). Cummins (1979) posited that the influence of cognitive load plays a significant factor in the achievement of ELLs, suggesting that developing competencies in L1 can help facilitate knowledge acquisition in L2 by reducing the cognitive capacity needed to consistently transfer information between two (or more) languages. Indeed, Wang (2017) found that growth in working memory positively mediated the relationship between ELL status (coded as monolingual English proficient, bilingual proficient, and bilingual limited English proficient) and mathematics growth, such that ELLs categorized as having lower ELP were predicted to grow more in working memory, which then positively predicted mathematics growth through third grade. Greenfader (2017) similarly used the ECLS-K:2011 and found that working memory was a stronger predictor of mathematics among Hispanic students than English-monolingual White students. Guglielmi (2012) provided evidence that the relationship between L1 in grades eight and 12 mathematics was mediated through L2 ability, suggesting cross-language transfer. Cárdenas-Hagan, Carlson, and Pollard-Durodola (2007) show this cross-language transfer in letter sound and identification at an earlier age (kindergarteners), although they did not observe that L1-L2 transfer occurred for phonological awareness and oral language among students instructed in English. Instructional context aside, Cárdenas-Hagan et al. and Guglielmi (2012) both showed, at varying developmental levels, that native language relates to developing second language abilities, which then go on to predict later achievement levels. Although neither study addressed the cognitive mechanism, this research further underscored the importance of L1 theoretically playing a cognitive role in acquiring L2. Learning difficulties in mathematics may compound this language and cognitive load problem by further straining short-term working memory and executive functioning.

Findings from meta-analyses (Grundy & Timmer, 2017; Linck et al., 2014) and recent longitudinal studies (Greenfader, 2017; Swanson, Kong, & Petcu, 2018; Wang, 2017) underscore the importance of understanding the role of executive functioning across language groups, especially working memory given the unique variance it consistently explains in developmental studies of mathematics. Students who are much lower in ELP may possess differential risk factors to lower mathematics performance over time, and it is possible language ability interacts with different dimensions of domain-general and domain-specific risk (poor mathematics performance). As such, it is important to describe the current knowledge of dimensions of risk for poor mathematics performance over time and the topography of mathematics skill development among at-risk learners.

Atypical Mathematics Development

Deficits in early numeracy skills (Chard et al., 2005; Jordan et al., 2006; Jordan, Glutting, Dyson, Hassinger-Das, & Irwin, 2012) and cognitive attributes such as processing speed (Moll et al., 2016) and working memory (Fuchs et al., 2005; Geary, 2011; Geary et al., 2017; Li & Geary, 2013; Moll et al., 2016) are risk factors for the development of significant mathematical difficulties. A multitude of studies have demonstrated the efficacy and effectiveness of early numeracy and number sense programs to prevent the compounding of mathematics difficulties over time (Codding et al., 2011; Fuchs et al., 2005; Jordan et al., 2006; Jordan et al., 2012). The Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (American Psychiatric Association, 2013) employs the term SLD with impairment in mathematics (i.e., mathematics SLD) to encompass a variety of barriers to typical mathematics skill development; the term dyscalculia was previously used to refer to this diagnosis and still is used frequently in the literature. These barriers may include problems related to mathematics facts fluency, computation, and numeracy (American Psychiatric Association, 2013). Prevalence of mathematics SLD ranges from 5 to 15% (American Psychological Association, 2013). Neurodevelopmental disorders such as Attention-Deficit/Hyperactivity Disorder (ADHD), SLD with impairment in reading as well as anxiety, and language deficits may be important secondary or co-occurring factors in mathematics SLD (Kaufman & von Aster, 2012).

A number of theories contribute to the current understanding of mathematics skills deficits. Landerl, Bevan, and Butterworth (2004) presented a number of working memory, numerical, and spatial-processing tasks to eight- and nine-year-old children with either reading SLD, mathematics SLD, or co-occurring reading and mathematics learning disorders. Although their study was more formative research and was ultimately unable to identify the relationships between working memory, verbal memory, numerical ability, and reading, they presented various findings associated with task-based deficits that children with mathematics SLD face. One significant finding was that children with mathematics SLD and reading plus mathematics SLD performed significantly slower than a control group and a reading SLD group on a number comparison task (no effect size reported). Similarly, children with mathematics SLD and reading plus mathematics SLD performed slower than controls and children with just reading SLD in a variety of number sequencing tasks (counting 45 to 65, 2 to 20, 20 backward to 1). Interestingly, among the contrasts conducted to follow-up with their ANCOVA analyses, no groups differed in counting from 1 to 20. However, no significant differences were detected between experimental groups on a dot comparison task involving counting a set of either 1-3 dots or 4-10 dots. This study significantly contributed to the theoretical evidence suggesting that impairments specifically relating to numeracy and quantity characterize mathematics learning difficulties. As the authors noted, their evidence corroborated findings from neuropsychology and cognitive neuroscience (Dehaene, Dehaene-Lambertz, & Cohen, 1998; Dehaene, Piazza, Pinel, & Cohen, 2003). Again, language proficiency was not controlled for or modeled as a mechanism of their observed findings, limiting these results in informing practice for linguistically diverse students.

Despite the evidence presented above, the cognitive neuroscience field faces a divide over the cognitive basis of mathematics learning difficulties. A number of studies (e.g., Andersson & Lyxell, 2007; Geary, Hoard, Nugent, & Bailey, 2012; McLean & Hitch, 1999; Szucs, Devine, Soltesz, Nobes, & Gabriel, 2013) have leveraged evidence against the magnitude processing camp, making a cogent argument that a large portion of problems in mathematics learning may be attributable to deficits in verbal, visual-spatial, and central executive components of working memory. Both verbal and visual-spatial working memory have been identified as components of mathematics learning disorders (Swanson & Zheng, 2014). Importantly, Szucs et al. (2013) identified visual-spatial working memory as one key deficit among individuals with mathematics SLD, leveraging evidence against studies suggesting a general magnitude processing perspective:

> It is noteworthy that Landerl et al. (2004) is one of the most often cited studies in support of the MR [magnitude representation] theory. However, that study merely demonstrated that DD have slower magnitude comparison speed than controls, which can happen for many reasons. (p. 2675)

The working memory argument, at a basic level, posits that impairments in numerical ability relate to impairments to accurately accessing learned information, manipulating that information, then systematically and efficiently carry out the task at hand. Thus, given the roles of phonological processing, language, as well as visual-spatial processing in early mathematics and in predicting mathematics trajectories (Li & Geary, 2013; Geary et al., 2017), deficits in verbal and visual-spatial processing are an unsurprising culprit in impeding mathematics learning. Swanson et al. (2018) extended such work to ELLs with mathematics difficulties, showing that growth in working memory was related to mathematics computation growth among English-Spanish proficient students. However, this was not the case among low-proficiency ELLs, indicating that the risk factors among low-proficiency ELLs with mathematics difficulties may take on a different profile than more English-proficient ELLs. Nevertheless, working memory played an important role in explaining variance in computation growth across all mathematics and language proficiency categories. Yet such findings must undergo corroboration in other studies given the varying definitions of ELLs in the research literature as well as in policy across states and school districts. Moreover, there is additional need to investigate these domain-general connections because working memory and language play meaningful role in mathematics development for typically and atypically developing children in mathematics and therefore represent key components in mathematics, seemingly regardless of skill level.

In general, these domain-specific and -general risk factors suggest that schools can address early number difficulties using multidimensional, focused screeners to identify and predict risk. As Clarke, Baker, Smolkowski, and Chard (2008) identified, kindergarten growth in skills measured on CBM, specifically quantity discrimination (e.g., discriminating between the larger of two numbers), predicted higher achievement on a standardized criterion measure. Additionally, static measurement of early numeracy via CBM as a whole has well-established concurrent and predictive validity (Clarke et al., 2014) for students at risk for mathematics difficulties. Growth on these componential numeracy skills is necessary to demonstrate adequate instructional response, and a lack thereof suggests the need for tiered intervention (Clarke et al., 2008; Fuchs et al., 2005). Later on, another risk factor in developing mathematics development is deficits in word problem solving (Fuchs, Powell, Cirino, et al., 2014) and fraction understanding (Resnick et al., 2016; Rinne, Ye, & Jordan, 2017). Given the sequence of mathematics skill development, with number sense leading to calculation ability which then forms the algorithmic basis for understanding word problems and pre-algebraic knowledge and concurrently fraction understanding, early identification and intervention is integral to prevention. And, especially given recent evidence that early intervention promotes students use of prior knowledge to increasingly acquire new mathematical knowledge (Watts et al., 2017), it is even more critical that students at risk for SLD in mathematics receive appropriately tiered instruction as, on top of content knowledge deficits, they may be less efficacious at recruiting and acquiring knowledge for problem solving given the deficits in executive functions (especially working memory) consistently observed among at-risk learners.

This further underscores the need to establish effective procedures and practices of RTI in mathematics for ELLs given the additional contextual, cognitive, and domain-specific variables that may impact their learning both positively and negatively. In essence, although literature has identified profiles of risk in mathematics learning, this infrequently accounts for students with limited ELP (with the exception of studies such as Davis & Powell, 2017; Doabler, Clarke, Kosty, et al., 2016, Doabler et al., 2019, and Swanson et al., 2018). Though similar patterns of deficits in mathematics knowledge may arise across language groups (e.g., poor number sense in kindergarten is likely to be a common risk factor regardless of language), such risk factors may be qualitatively distinct, requiring researchers and practitioners to attend to the qualities of learning factors (e.g., deficits in multiple areas processing speed, language, working memory, level of mathematics knowledge in different areas) just as much as the degree of the deficit in grade-level performance.

An important aspect of this literature to note is that few studies investigate the potential unobserved patterns of development among ELLs, though a handful of studies address unobserved trajectories in number sense (Jordan et al., 2006; Jordan et al., 2007) and fraction knowledge development (Resnick et al., 2016; Rinne et al., 2017). These prior studies are highly informative in establishing a general sense of what constitutes “at risk” developmental patterns, though it is necessary to understand the profile and distribution of growth trends specifically within ELLs to explore if these ELL patterns match the more general profiles of mathematics growth mixture distributions. That said, the detection of these is not directly related to identifying of mathematics difficulties or learning disorders. What they are informative for, however, is distinguishing among qualitatively distinct development profiles that are suggestive of at risk developmental patterns. Detection of such profiles based on those qualitative factors across grades may then be mapped onto current knowledge of how mathematics difficulties and learning disorders function over early and intermediate elementary school, informing assessment and intervention practices.

Types of Mathematical Knowledge

Although the focus of much of the literature review up to this point has been focused on explanatory mechanisms of achievement gaps between at-risk math learners and typically developing ELLs and EPSs, it is also important to take a broader, developmental perspective on how learners acquire mathematics skills. This is important not only to broadly contextualize what predicts math achievement at a cognitive level, but it provides insight into the ways in which knowledge and skill acquisition interact with factors that are exogenous and endogenous to the student.

Acquiring mathematics skills and knowledge requires both conceptual and procedural knowledge (Rittle-Johnson, Siegler, & Alibali, 2001). As Rittle-Johnson et al. (2001) discussed, there existed “concepts-first” (Byrnes & Wasik 1991; Byrnes, 1992) or “procedures-first” (Siegler, 1991) theories, which argued that concepts or procedures preceded the development of the other skill. Although experimental evidence seemed to substantiate both theories, Rittle-Johnson et al. (2001) provided causal evidence that these processes and skills reciprocally and simultaneously impact each other, and that the representation of the problem in working memory mediates the development and reciprocity of the two. However, research over the past decade or so has yet to settle on definite properties of either concept, making the measurement of either domain of mathematical knowledge challenging (Baroody, Feil, & Johnson, 2007; Crooks &

Alibali, 2014). That said, Crooks and Alibali (2014) noted that the momentum of recent policy, curricula, and research is driven towards building conceptual understanding, particularly as it relates to indirectly building procedural knowledge. Baroody et al. (2007) made a convincing case for this perspective, as they proposed a model for integrating conceptual and procedural domains. Paraphrasing the authors, they discussed that procedural knowledge acquired in isolation of concepts (or with superficial connections to concepts) may lead to more inaccurate problem solving or strategy use (Baroody et al., 2007, p. 126). Recent tier-2 intervention approaches in kindergarten build on the reciprocal nature of concepts and procedures (Clarke, Doabler, Smolkowski, Baker, et al., 2016; Clarke, Dobaler, Smolkowski, Nelson, et al., 2016), both of which are often deficits in students with or at risk of specific learning disorder in mathematics. Results of randomized trials indicated immediate, but not long-term, efficacy of the ROOTS curriculum across a number of fluency- and conceptually-based assessments (Clarke, Doabler, Smolkowski, Baker, et al., 2016; Clarke, Doabler, Smolkowski, Nelson, et al., 2016).

Although the studies above provide concrete evidence for these conceptual and procedural processes in typical and at-risk individuals, it is still unclear as to how these concepts generalize to ELLs, for whom the addition of language may mediate and/or moderate mechanisms of knowledge and skill acquisition. As an example, Rittle-Johson et al. (2001) stated that representations of knowledge are critical in relating conceptual and procedural understanding. However, the added cognitive load of language barriers may impact how these representations are generated and utilized. If the student is at risk for mathematical difficulties (independent of any language barriers), language and executive-functioning deficits could potentially cascade to severely limit a student’s ability to meaningfully create representations of information needed to connect the underlying concept of a problem with the strategy used to execute it. Additionally, these processes may differ significantly across mathematical domains. The mechanisms of developing procedural and conceptual understanding may differ more in kindergarten than in third grade for ELLs, as the language demands might be much more challenging; learning cardinality and quantity-symbol correspondence assumedly differs significantly in the language needed from prealgebraic concepts, fractions, and word problems. Although evidence suggests that language could be important for responsiveness to mathematics interventions across many student demographics and skill levels (Doabler, Clarke, Kosty, et al., 2016; Doabler et al., 2019 Powell et al., 2017), problem representation, as well as the ways in which procedural and conceptual understanding reciprocate, may appear different for students navigating second language learning and barriers. Even if students are assessed in their knowledge in their native language, instruction likely is not fully in their native language, at times requiring students to transfer knowledge learned in or translated from a second language back into a native language to more fully understand the tasks and concepts at hand.

Instructional Factors

Although computation itself is not inherently language based (Gelman & Butterworth, 2005), the mediums through which mathematics is communicated (oral, written, symbolic) are. On an instructional basis, mathematics is laden with academic language (Albers et al., 2013; Kersaint et al., 2013). One major component of effective instruction for ELLs is developing this unique area of academic language (Gersten et al., 2007). Relatedly, mathematics has a linguistic "'register'" composed of four different elements: vocabulary, syntax, semantics, and discourse (Kersaint et al., 2013, p. 43). For example, regardless of whether actual mathematical computation utilizes language, teachers use content-specific vocabulary to teach the conceptual and problem-solving skills needed for the arithmetic. This content-specific, academic language only increases as students age and the conceptual and procedural bases of mathematics become increasingly computationally complex. For example, it requires even more linguistic demand for a student to extract vocabulary information from the general syntax of a word problem, define the conditions of the word problem, and finally synthesize that information to form a coherent algorithm to complete the problem at hand (Kersaint et al., 2013). Underlying these linguistic demands are, of course, the student's mathematical knowledge needed to solve the problem. On an instructional level, the explanations a teacher may use to teach addition, subtraction, multiplication, and division may be confounded and obscured by a students' language proficiency ("oral discourse"; Kersaint et al., 2013). A simple explanation of the number line, subtraction, or division may not be easily understood without making accommodations to meet the needs of proficiency levels or adjusting how the content is presented. Jayanthi, Gersten, and Baker (2008) advocated for the use of explicit instruction techniques for at-risk mathematics learners, many strategies of which emphasize the opportunities to verbalize problem solving skills. Doabler, Nelson, and Clarke (2016) further emphasized the components of explicit instruction, especially problem verbalization, as a key mechanism to helping ELLs at risk for mathematics difficulties build their mathematical knowledge component. These instructional factors require students to have an adequate level of English proficiency to understand direct explanations of mathematical concepts in addition to synonyms and alternative explanations of those concepts, and techniques such as explicit instruction can help leverage ELLs' nascent L1 and L2 proficiencies to access rigorous mathematics content. Doabler, Clarke, Kosty, et al. (2016) conducted secondary analyses of an RCT of a universal kindergarten mathematics curriculum specifically looking at the impact on ELLs' mathematics achievement. Their findings indicated that the curriculum was effective for ELLs (as it was for EPSs; Clarke et al., 2011), and prior low performance in mathematics did not moderate these effects. As the authors discuss, this did not close early achievement gaps between lower and higher performing ELLs based on pretest scores, though the curriculum effects are indicative of the role of language and instructional discourse in early mathematics learning.

Much of the curriculum and instruction research on this area focuses on the instructional components that best lead to student success (Loeb et al., 2014; Master et al., 2012), although little of that inquiry has focused on the mechanisms contributing to differential impacts on ELLs versus EPSs. Some research, though, accomplishes this through qualitative methods. Gutiérrez (2002) conducted a qualitative analysis of three high school mathematics teachers serving ELLs in a mixed Spanish and English classroom based on classroom observations, teacher interviews, and student interviews. Gutiérrez distilled these strategies into a number of important characteristics, such as scaffolding prior knowledge and allowing students to use their primary language as necessary (especially with particularly cognitively demanding tasks). More broadly, Gutiérrez also expanded on previous research by noting the importance of teacher understanding and empathy of linguistic background and diversity. As aforementioned, additional teacher-level factors including training, professional development, and experience with ELLs are important in garnering ELLs' learning strengths (Loeb et al., 2014; Master et al., 2012).

Tiered intervention effects for at-risk mathematics learners. Currently, there is a dearth in research in the area of identifying risk for math-related difficulties in ELLs, although assessment of mathematics among ELLs through summative measures has been well documented (Abedi, 2002; Abedi & Lord, 2001). Recently, research has employed more prevention-based techniques to address mathematics difficulty among kindergarten-aged ELLs

(Doabler, Clarke, Kosty, et al., 2016). Although this research focuses on addressing mathematics difficulty, it still does not propose a better system by which these students can be identified. Intuitively, the earlier students can be identified, the better. Operating within a prevention model is a major component of RTI; however, Doabler, Clarke, Kosty, et al. (2016) found that the curriculum that was under investigation (i.e., Early Learning in Mathematics [ELM]) produced equally positive effects for students across skill levels (Hedge's *g*: Test of Early Mathematics Ability-Third Edition [TEMA-3] = 0.30 with statistically significant effect, Early Mathematics Curriculum-Based Measure = 0.18 without statistically significant effect). This finding is important in terms of the efficacy of this curriculum for kindergarten ELLs, yet the authors highlighted that these findings were contradictory to a previous study (Clarke et al., 2011) that investigated the effects of ELM for students with mathematics learning difficulties, which found that ELM yielded stronger effects for students with lower prior mathematics achievement. Doabler, Clarke, Kosty et al. rationalized their effect by noting that the explicit instruction components of ELM may be the important mechanism in helping ELLs and students with mathematics learning difficulties (and ELLs with learning difficulties) learn. As a universal technique to prevent mathematics learning difficulties at an early age among ELLs, this study has significant implications that are greatly enhanced by the methodological rigor of their randomized trial. However, fitting this curriculum into the context of screening measures may make it difficult to accurately identify the students considered at risk for mathematics difficulties if the effects are uniform across ability groups. These uniform effects are undoubtedly positive among the ELL population as it would seem to suggest that mathematics learning difficulties and limited English proficiency can be equally balanced (instructionally) using techniques of explicit teaching. Additionally, these results seem to suggest that the language of mathematics is important for all students, not just those considered at risk.

However, in a study of a tier-2 mathematics intervention, Doabler et al. (2019) found that ELLs' ELP level did not significantly moderate the intervention effect at the .05 level, though the authors highlighted the potentially meaningful variation in the treatment effect across ELP scores may suggest at-risk ELLs with low ELP responded meaningfully stronger than those with higher ELP. Similar to Doabler, Clarke, Kosty, et al. (2016), though, prior mathematics ability did not moderate treatment effects among ELLs. Such findings further underscore the role that language could play in treatment response and how it serves as an additional contributor to how mathematical knowledge is acquired and utilized. To that end, language and the role of other domain-general abilities require further investigation to better understand factors that may contribute to response to instruction.

The Present Studies

Many studies establish cogent correlational evidence for these language-mathematics relationships among ELLs (Chen & Chaloulb-Deville, 2016; Garrett, 2010; Greenfader, 2017; Roberts & Bryant, 2011; Wang, 2017); some studies even establish a causal relationship between language and mathematics (Hassinger-Das et al., 2015; Pupura et al., 2017). However, integrating across many different disciplines and drawing on developmental, cognitive, assessment, and intervention studies, it is apparent that there is a significant lack of understanding regarding the patterns of mathematics development among ELLs, the domain-general predictors (i.e., language and working memory) of those trajectories, and for whom these patterns may differentially unfold. Additionally, despite the increasing work in this area, few studies extended this line of inquiry specifically to the RTI framework in which a prevention focus is explicit in identifying student risk status (Burns, 2011). Designing optimal prevention techniques requires an understanding of the developmental trajectory underlying skill development, stability, change within that developmental process. Moreover, predictors of those developmental trends become critical to the provision of services to support academic, cognitive, and behavioral development. Of those predictors, language is arguably the most critical factor among ELLs to investigate as a key leverage point in targeting mathematics intervention due to its obvious but nuanced and relatively understudied role in mathematics development. Understanding the relationship of ELP to developmental trends, however, also requires understanding the prior-established cognitive components that predict mathematics development. Working memory consistently emerges as a key player.

The current studies merge these concepts in the following manner. Evidence suggesting that language is critical to understanding mathematics (Vukovic & Lesaux, 2013), execution of different problem types (Spelke & Tviskin, 2001; Venkatraman et al., 2006), and progressing adequately with peers (Doabler, Clarke, Kosty, et al., 2016; Doabler et al., 2019; Halle et al., 2012; Roberts & Bryant, 2011) can be merged with theories on cognitive load (Cummins, 1979; Sweller et al., 2011), mathematics knowledge types (i.e., conceptual, procedural; Baroody et al., 2007; Crooks & Alibali, 2014; Rittle-Johnson & Alibali, 1999; Rittle-Johnson et al., 2001), and the general importance of mathematics for academic development (Duncan et al., 2007) to provide a framework for understanding potential mechanisms of how language and mathematics are related. From a broader developmental perspective, the interactions of second language proficiency, general language, mathematics, and the ecology of schools necessitates a developmental systems and bioecological framework to contextualize growth in mathematics given the cascading relationships between social and educational context, cognitive, social, emotional, behavioral, and academic functioning. Language may be a critical mechanism for generating understanding of mathematical concepts (Vukovic & Lesaux, 2013), although language proficiency could be related to the rate at which both conceptual and procedural understanding (as a reciprocal relationship) are acquired, thereby accounting for the amount of change or gains students make in mathematics and, in turn, how children understand mathematics. Thus, the long-term interdependent cycle of procedural and conceptual knowledge types building on each other may be mitigated (or accelerated) if language partially accounts for how students understand mathematics (e.g., Vukovic & Lesaux, 2013). Adding to this, executive functioning ability specifically has been implicated in the acquisition of specific types of mathematics knowledge (Cragg & Gillmore, 2014; Cragg et al., 2017) as well as bilingualism (Bialystok, 2011). Thus, it is possible that executive functioning abilities may moderate how ELP is associated with mathematics by playing a role in how mathematical knowledge is acquiried and implemented, which could then predict trends mathematics growth over time. Complementarily, working memory may function differentially among ELLs and EPSs given the varying patterns of domain-general and domain-specific skills development. A logic model depicting these long-term relationships between conceptual and procedural knowledge, ELP, and cognitive attributes is presented in Figure 1 below. Although not all aspects (i.e., conceptual and procedural knowledge) are directly measured in the current data, they are assumed to be inherently implicated in the development of mathematical knowledge given the current instructional emphases and content of assessments in mathematics. The unmeasured aspects of this model can serve as the basis for further empirical and theoretical investigation for ELLs and EPSs alike, as language serves as a general predictor of mathematics performance regardless of second language learning (LeFevre et al., 2010; Powell et al., 2017; Vukovic & Lesaux, 2013).

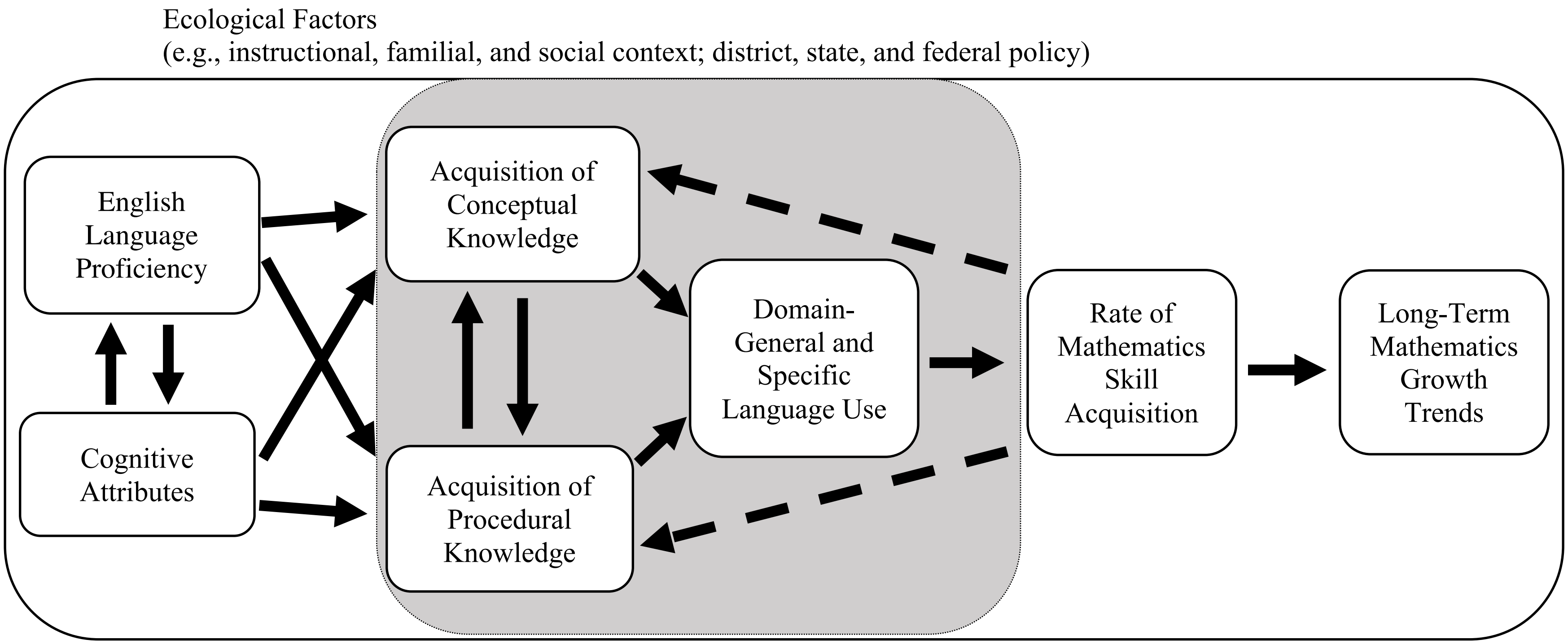

Figure 1. Hypothesized logic model of mathematics growth and change over time for ELLs.

Note. Dotted arrows indicate a feedback loop, whereby acquiring skills faster allows an individual to build conceptual and procedural knowledge quicker, though this growth diminishes over time for some students. Information not included in the gray box is the subject of empirical investigation in the present studies.

The present investigation is divided into two studies. The first study assesses differences in change patterns across time between ELLs and EPSs. Much of the research related to mathematics achievement among ELLs compares achievement between ELLs and their EPS peers. This study seeks to extend that line of research by employing novel forms of change modeling to better understand the nature of interindividual differences in intraindividual change in addition to offering a more practice-oriented approach to considering ELL characteristics. In Study 1, I employ multiple group latent change score modeling (MG-LCSM; McArdle & Nesselroade, 2014) which allows me to analyze differences in growth patterns through fixing equality constraints across groups to test if the model is better fit without those equality constraints. In doing so, I can directly test the differences in growth parameters across groups while modeling the change within each group separately.

Given the prior research on the role working memory may play across language groups (Greenfader, 2017; Swanson et al., 2018; Wang, 2017) and its relationship to mathematics, I also investigate the extent to which working memory may differentially predict mathematics development patterns net of other dimensions of executive functioning (i.e., inhibition and cognitive flexibility). Both Greenfader (2017) and Wang (2017) addressed the role of working memory as mediators between language group and later outcomes (although Greenfader [2017] also examined how language group moderated working memory). However, Study 1 addresses whether working memory skills at the end of kindergarten differentially predict mathematics development patterns of through grade four.

In Study 2, I exploratorily investigate heterogenous growth patterns utilizing growth mixture modeling (GMM) as well as predictors of heterogenous mathematics development, specifically growth in English language proficiency in kindergarten, working memory, and the interaction between ELP gains and working memory. Utilizing data from Spring of first grade through Spring of fourth grade, I test the patterns of growth among students classified as ELLs in kindergarten once the majority of these students are fully assessed on their mathematics skills in English. The primary goal of this study is to detect which students may potentially be classified as at risk based on their initial level in their growth trajectory and the ensuing pattern of growth. Prior research has exhibited the ways in which the GMM technique is useful for identifying students who are stable in their trajectories over time or exhibit differential patterns of growth (Kaplan, 2002). Students who exhibit more stable, low-growth patterns may indicate academic risk by not acquiring skills at a normative pace to equip them with the necessary skills for long-term mathematics success, especially as students transition into fraction and algebra learning. However, stability in membership to growth trajectories can also be tested in mixture modeling. These models, called latent markov chain models (Kaplan, 2008) or regime switching models within a growth model (Chow et al., 2013), are particularly useful to investigate questions of discontinuity in development, provided heterogenous patterns of growth exist. Overall, Study 2 investigates the extent to which the interaction of early language proficiency growth and working memory predict growth in mathematics trajectories in middle elementary school and stability in that growth in an effort to unpack the role of working memory in the relationship between English language proficiency and mathematics achievement trends.

CHAPTER III: STUDY 1

Research Questions and Hypotheses

Research Question 1

Does the pattern of mathematics growth from the Spring of kindergarten to the Spring of grade 4 differ significantly for ELLs and EPSs in terms of additive and proportional change? Research utilizing ECLS-K: 1999 data showed that ELLs proficient in English by kindergarten entry exhibited steeper growth in mathematics relative to EPS peers through grade eight (Halle et al., 2012). However, these studies (in both conditional and unconditional models) employed a linear slope in their growth model. Multiple studies have exhibited the decelerating of academic growth over time, often in a piecewise (Kohli, Sullivan, Sadeh, & Zopluoglu, 2015), nonlinear (Grimm, Ram, & Estabrook, 2010), or quadratic (Roberts & Bryant, 2011) manner. Linear models of growth, such as Halle et al. (2012) who showed English-proficient (at kindergarten entry) ELLs exhibited steeper growth through grade eight in math compared to EPS peers, potentially omit meaningful patterns of variability across development. A similar story can be told with Garrett’s (2010) and Wang’s (2017) results, in which linear trajectories were favored as the primary functional form of development through grades eight and three, respectively. In the current study, I hypothesized that the intercept as well as constant and proportional change components of the MG-LCSM will differ significantly between groups. Specifically, I predicted that ELLs will be significantly lower on their intercept, exhibit steeper growth than EPSs over early elementary, then – as has been identified in previous literature – decelerate in academic growth toward the end of intermediate elementary. I predicted that ELLs, relative to earlier change, will decelerate more than EPS peers, which may potentially be attributable to the compounding of disadvantaged mathematics performance early on coupled with increasing linguistic complexity of mathematics. Recent work has shown the importance of recruiting prior knowledge to continually building mathematics knowledge ("state" levels) and enhance performance over and above "trait" levels, i.e., the portion of individual differences in mathematics performance level that remains stable over time (Watts et al., 2017). Thus, given the multidimensional and sequential nature of mathematics development, early performance deficits portend poorer knowledge-building mechanisms, disadvantaging students' capacity to recruit prior knowledge as building blocks for new mathematics concepts and procedures. Early language barriers among less English-proficient ELLs may further inhibit the knowledge building and application process. Compared to EPSs, these language and mathematics development factors together create more barriers to bridging early numeracy and arithmetic knowledge from early elementary to later algebraic thinking, fraction learning, and word problem solving, thereby differentially slowing performance gains on average.

Research Question 2

To what extent does working memory differ across ELLs and EPSs in predicting mathematics development patterns? This question is largely exploratory in nature given the limited research on the contribution of early working memory to mathematics development among lower-ELP ELLs versus EPSs, though recent work has shown that working memory is more strongly predictive of second grade mathematics achievement among Hispanic students and Hispanic students with lower early literacy performance (Greenfader, 2017). The significant prior research on the relationship between working memory and mathematics among both typically developing (Geary et al., 2017**;** Lee & Bull 2016**),** at-risk children (Fuchs, Schumacher, et al., 2014; Swanson & Zheng, 2014), and ELLs (Swanson et al. , 2018) suggests that working memory plays a substantial role in explaining variation in mathematics achievement concurrently and longitudinally. However, because language has been consistently identified to be related to mathematics development, mathematics tasks may recruit memory resources in a different fashion over time for ELLs given the additional language load of mathematics tasks among ELLs. In particular, early individual differences in working memory capacity may be differentially predictive of later mathematics development trends given this potential the relationship between language and working memory to compound over time in relation to mathematics.

Study 1 Method

In Study 1, I utilized data from the Early Childhood Longitudinal Dataset: 2010-2011 Kindergarten Cohort Public-Use data file (ECLS-K:2011). These data are available from the National Center for Education Statistics website. The ECLS-K employed a three-stage complex sampling design. In stage one, 90 primary sampling units (PSUs) comprised of counties were selected, followed by schools within PSUs in stage two, and finally students within schools in stage three . In the kindergarten base year, study staff collected data from 18,174 students on a wide range of measures assessed at the student, parent, teacher, administrator, and school levels. These students were followed through the end of grade five, with 30% random samples taken in the Fall of first and second grade. Students in the full ECLS-K: 2011 sample attended both private and public schools. Trained study personnel administered direct child assessments. At the time of conducting the current research, only data through grade four were publicly available.

Participants

In Study 1, I utilized data only from students who attended public schools in the fall and spring of kindergarten and the spring of first through fourth grade. I used only public school students predominantly because the role of RTI (or similar models) in private schools is likely highly variable (or nonexistent) and to remain consistent with prior analyses of academic development of ELLs (Roberts & Bryant, 2011). Thus, the context of public schools is an important factor in this study. Roberts and Bryant (2011) took a similar approach, additionally justifying that the context of private schools substantially different than public schools. I defined ELLs as the students whose parents reported that English was *not* the primary language at home (in kindergarten) and who were not fully English language proficiency in the Fall of kindergarten (discussed more below).

Measures

Mathematics. The mathematics achievement measure in the ECLS-K: 2011 is a vertically-scaled, item-response theory (IRT)-based measure (Tourangeau et al., 2018). Narjarian et al. (2018b) reported that differential item functioning (DIF) tests revealed no significant impact of language on the item characteristics, suggesting that the same mathematics skill on each item was being assessed across Spanish and English administrations. Students who were assessed in Spanish are discussed more below. The mathematics IRT measure has an α estimate of between .92 and .94 across all rounds of assessment in the full sample.

English Basic Reading Skills. All students, regardless English proficiency, received part of the IRT reading measure in English called the English Basic Reading Skills (EBRS) measure. Twenty items comprised this measure: 18 items on basic English literacy skills (e.g., letter-word identification, phonemic awareness) as well as two items from the Preschool Language Assessment Scale (*pre*LAS) Art Show subset, which measures expressive vocabulary using pictures (Tourangeau et al., 2015a). Because all students received the EBRS, all students also could obtain a score on the ECLS reading IRT measure, regardless of their actual routing through the multi-stage IRT assessment (Tourangeau et al., 2015a). However, nine students who were routed through the cognitive battery in Spanish did not have an IRT score, though they obtained a score on the EBRS of either 0 or 1. As such, the EBRS was used as a covariate in all Study 1 analyses rather than the reading IRT measure, though students' raw EBRS score does not reflect IRT dimensions such as item difficulty or the same content coverage of the larger reading assessment (e.g., reading comprehension; Tourangeau et al., 2015a). Nevertheless, the EBRS data helps control for early native and second language reading (as opposed to oral language proficiency) abilities across and within groups. For Fall 2010, the EBRS had a researcher-reported α of .87 (Tourangeau et al., 2015a).

Socioeconomic status. Socioeconomic status (SES) measured at baseline (i.e., 2010-2011) is a standardized (M = 0, SD = 1) composite measure comprised of parent occupational prestige, income, and education level data (Tourangeau et al., 2015a).

Pre-kindergarten care arrangement. Pre-kindergarten care arrangement was a categorical variable with two levels indicating the type of care arrangements children received prior to entering kindergarten. Parents provided these data on the kindergarten parent questionnaire (Tourangeau et al., 2015a). This variable was turned in to a binary indicator variable with values related to whether the child received parental care only (1) or nonparental care (0) during the year prior to kindergarten. The majority of students with nonparental care arrangements attended center-based care programs (67%), with relative care in the child's or relative's home the next most frequent form of nonparental care (19%).

Kindergarten type. Kindergarten type was a dummy-coded variable indicating whether the child attending half or full day kindergarten in the Fall of 2010 (1 = *full day*, 0 = *half day*; Tourangeau et al., 2015a). In half-day kindergarten, students attended morning or afternoon classes; however, for the purposes of this study, I do not make a distinction between morning and afternoon half-day kindergarten.

First-time kindergarten status. Students in the kindergarten base-year sample of the ECLS-K: 2011 had data collected regarding whether they had previously attended kindergarten (Tourangeau et al., 2015a). This is a dummy-coded variable indicating whether the student had previously attended kindergarten (1 = *yes*) or not (0 = *no*). Many prior studies have focused analyses only on first-time kindergarteners; however, because the focus of this study was not related specifically to aspects of kindergarten entry of effects of kindergarten, first-time kindergarten status was used as a control rather than an exclusionary category.

Age at kindergarten entry. Age at kindergarten entry was measured in months (Tourangeau et al., 2015a). Age at kindergarten entry measured age of students when they entered kindergarten for the first time and would likely help remove differences in individual and group differences that would be due to differential levels of mathematics development at kindergarten entry related to age.

Sex. Children's sex was obtained from school records and parent-report (Tourangeau et al., 2018). This variable was dummy-coded (0 = *male*, 1 = *female*).

Race/ethnicity. Race/ethnicity was obtained in the ECLS-K: 2011 data from parent report and school records (Tourangeau et al., 2018). Eight categories were provided as response options. These categories were transformed into eight dummy-codes to reflect each race category. However, due to some very small cell sizes, especially among the students defined as ELLs, a number of categories were collapsed into a single race/ethnicity variable for the multiple-group models. These race/ethnicity categories included students whose race/ethnicity was reported as Black/African American, American Indian, Native Hawaiian/Pacific Islander, Asian, and two or more races. This is predominantly a necessity of the modeling technique (described more below). Because race/ethnicity is utilized as covariates in each group of the multiple-group model, regressing the slope and intercept of the change score model onto very small *n* or zero-cell race/ethnicity categories may become problematic. By collapsing certain small-sample race categories into one, the same dummy codes can effectively be utilized in regressions within each group of the multiple-group model. This may underestimate variance that can be distinctively attributed to certain racial/ethnic categories; however, the focus is accounting for race/ethnicity variance broadly as opposed to modeling group-specific contributions to the model. Additionally, given the that the multiple-group aspect of the model is based on ELL-status is strongly confounded with race/ethnicity (i.e., the majority of ELLs in these data are of Hispanic descent), the multiple-group structure of the model is inherently accounting for a significant portion of the race/ethnicity variance in the outcomes of interest. The goal using the current race/ethnicity categories is to account for variance within each group. This is an important strategy as well, particularly among the ELL group. Roberts and Bryant (2011) utilized the ECLS-K: 1999 cohort to analyze developmental trajectories in mathematics between native Spanish-speaking and native Asian-language speaking ELLs, finding significant differences between the two groups over time. It is not the goal of the current study to model these differences in native language, although it is important to control for these differences in light of these prior findings.

English language proficiency. English language proficiency was assessed on the preschool Language Acquisition Screener (*pre*LAS; Duncan & De Avila, 1998). This measure is comprised of two oral language tasks, each with 10 possible correct responses. The entire screener has a total raw score of 20, although the "Simon Says" task is weighted twice as much as the “Art Show” task (Tourangeau et al., 2018). The total weighted score is 30. The raw number-right score (i.e., 0-20) for Fall 2010 has a α of .91 (Tourangeau et al., 2015a). Students who did not score at least 16 out of 30 on the screener were deemed not English proficient for the purposes of assessment. Those students who scored below 16 out of 30 on the *pre*LAS and spoke Spanish were administered all cognitive assessments in Spanish besides the English Basic Reading Skills measure. It is also important to note that the manner by which students were identified as ELLs in the current study does not align with how the ECLS-K:2011 researchers identified students who needed to take the Spanish version of the assessments or be excluded from assessment (e.g., some students not identified as ELLs still took assessments in Spanish). The definition of ELL in the current study was based on two factors: parent-report of home language in kindergarten (not English, English, or unable to choose primary language) and a score on the *pre*LAS. The ECLS-K: 2011 researchers used a publisher-recommended cutoff score of 16 to identify students who to take direct cognitive assessment in Spanish (among those who spoke Spanish). However, this cutoff score represents a narrow subset of the non-native English speakers in the dataset who are highly limited in their English proficiency. There is a large range to gaining English language proficiency that is not represented by utilizing such a low cutoff. State accountability and growth assessments such as the ACCESS for ELLs measure (World-Class Instructional Design, 2014) assess English proficiency along a broad continuum, from not English-speaking at all to highly English proficient. Thus, a higher cutoff was used to capture the larger variability in English language proficiency among ELLs that was more generalizable to the range of English proficiency found in a typical school district.

To establish this cutoff, the distribution of the *pre*LAS measure among students whose parents reported the home language was English or non-English (in kindergarten) was investigated to extract an appropriate cut score that would adequately distinguish between students with high ELP and students who exhibited lower (but not only very low) ELP .[1] The investigation into this cut score focused on two main elements: the extent to which the cut score distinguishes between EPSs who took the *pre*LAS and students whose home language was not reported as English, and the distinguishability of the cut score within the non-English home language group from ELP and limited ELP . To establish these proficiency levels, I examined the distribution of Fall 2010 *pre*LAS scores in the full kindergarten base year sample (N = 12,425) using the sample weight WC_2P_2TZ0 with the corresponding primary sampling unit (WC_2P_2TZPSU) and stratum (WC_2P_2TZSTR) variables. Table 3 displays the percentages of students within each band of English proficiency (I defined the bands for the purposes of this table). A Pearson χ^2 test with Rao and Scott adjustment in the *survey* package of *R* (Lumley, 2004; 2018) shows that the proportions of each category between the two groups are significantly (F = 541.66, numerator df = 2.77, denominator df = 443.31, $p < .05$).

[1] The *pre*LASscreening measure in the ECLS-K: 2011 data consists of a receptive (Simon Says) and an expressive (Art Show) language task. The Simons Says task score is weighted twice as much as the Art Show task. The score on the measure was generated using the formula described in the ECLS – K: 2011 K-4 User Manual: *pre*LAS score = Art Show + (Simon Says*2). The total weighted score on this measure was 30. Refer to the K-4 User Manual (Tourangeau et al., 2018) for more information.

Table 2
Percentages of Students in Language Proficiency Categories

Language Proficiency Level	Speaks English as Primary Home Language: Yes	Speaks English as Primary Home Language: No
English Proficient (≥ 28)	86.01% (85.91%)	33.05% (35.05%)
Marginal English Proficiency (<28)	11.25% (11.29%)	29.34% (29.23%)
Limited English Proficiency (<23)	1.87% (1.90%)	17.42% (17.41%)
Very Low English Proficiency (<17)	0.88% (0.89%)	20.18% (18.25%)

Note. Unweighted percentages in parentheses.

Figure 2 displays box plots and raw data points for *pre*LAS scores by parent-reported primary language spoken at home (which was recorded in either the Fall or Spring of kindergarten). Figure 3 displays the weighted boxplot using base-year sample weights that adjusted for student, parent, teacher, and before and after school program nonresponse. Comparing Figures 2 and 3, the quantiles of the unweighted data reflect similar distributions in the weighted data, though the boxplot whiskers for the weighted non-English speaking data are slightly different. Examining Figure 1, clusters of scores emerge in the data, where the overwhelming majority of EPS students score at 26 or above. Meanwhile, the median score for non-native English-speaking students is 25 (which is the lower whisker bound for EPS students). At and below a score of 22, there are proportionally few EPSs, while there is a relatively even distribution of students who spoke a non-English language at home between a score of 0 and 22. Given this breakdown of *pre*LAS scores, the cut-point for defining ELLs was set at a score of 23. This represents what is thought to be a conservative estimate of the full range of English proficiency, given that assessments such as ACCESS and other state-level proficiency measures assess students who are likely English proficient, while also taking into account a large profile of language domains (i.e., reading, listening, speaking, and writing). Based on this distribution, a cut score of 23 appeared to clearly distinguish ELLs who are limited in English proficiency relative to native English-speaking students (or students' whose parents reported speaking two primary home languages). In terms of the raw *pre*LAS total score, the subpopulation of 303 students who obtained weighted score of 22 or below obtained a maximum raw score of 16. Previous studies of ELLs using the ECLS-K:2011 utilized a raw cut-score of 16 to define ELL (e.g., Hartano et al., 2018; Wang, 2017; Wilkinson, 2017). Despite the raw score cut point's (16 out of 20) similarity to the weighted cut-score in the current data, using the publisher-recommended weighting of the *pre*LAS did not result in the same sample that would have been obtained using the total raw score because some students would have scored higher on the weighted metric (i.e., total score of 30). Using a raw cut-score of 16 out of 20 resulted in a sample size of 350. To adhere to the scoring protocol ECLS-K: 2011 researchers used to identify students who were to be assessed in Spanish, however, the weighted cut score of 23 was retained.

That said, this attempt to capture a level of English proficiency beyond what has been typically represented in prior work by taking more explicit effort to define English proficiency in a nationally-representative sample of the kindergarten base year is significantly limited due to the simple fact that the *pre*LAS is not nearly as comprehensive as the ELP assessments employed in schools. In practice, the definition of ELLs employed in schools varies significantly across states; there is no objective, absolute level at which English proficiency is determined. To that end, a relative proficiency approach such as this potentially reflects a more representative method of what constitutes English proficiency in school settings.

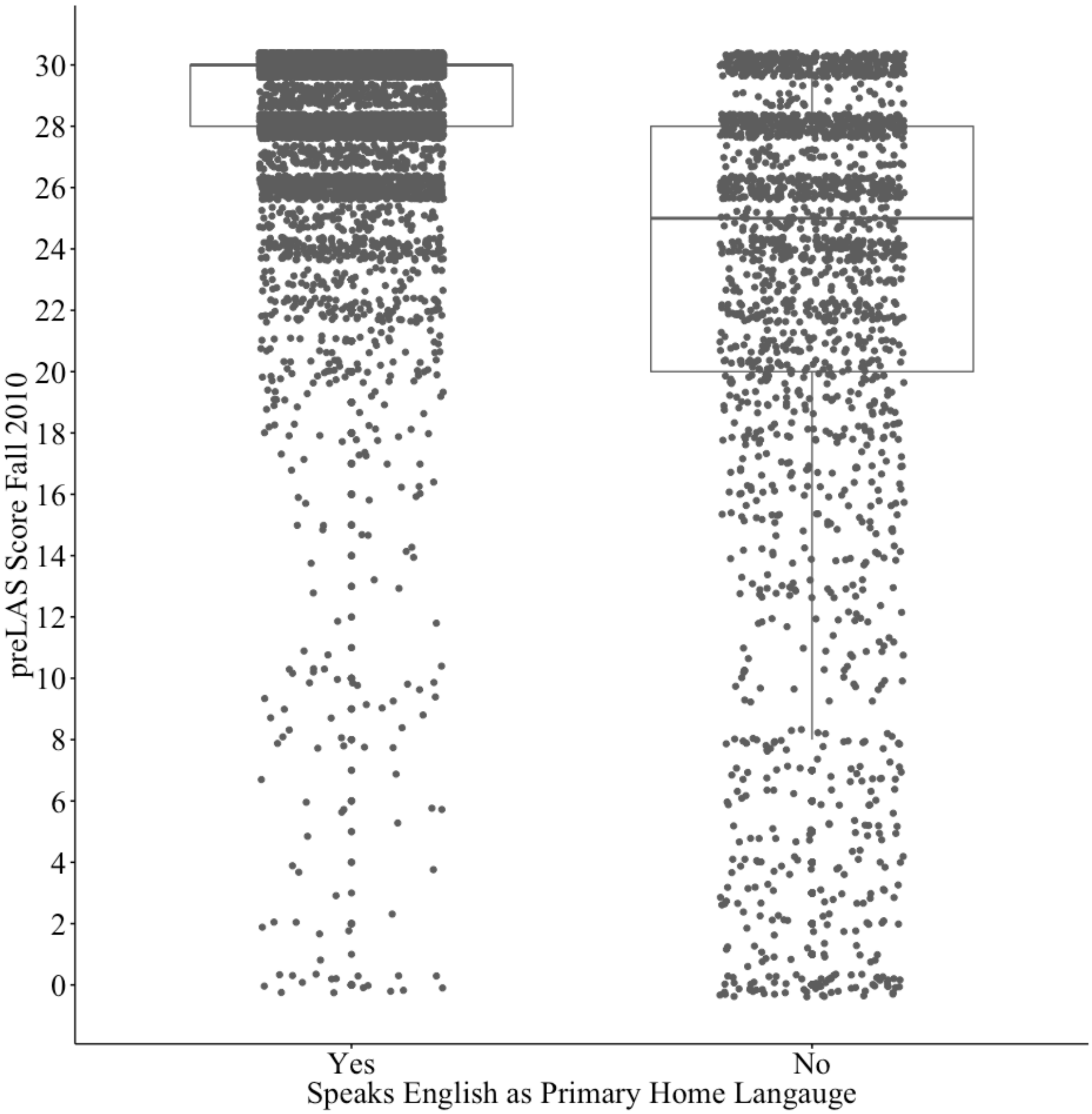

Figure 2. Unweighted boxplots and raw data points of Fall 2010 kindergarten sample *pre*LAS scores by primary home language.

Note. Parents of students who reported more than one primary home language were included in "Yes" for the purposes of this study. Figure produced in *ggplot2* (Wickham, 2016).

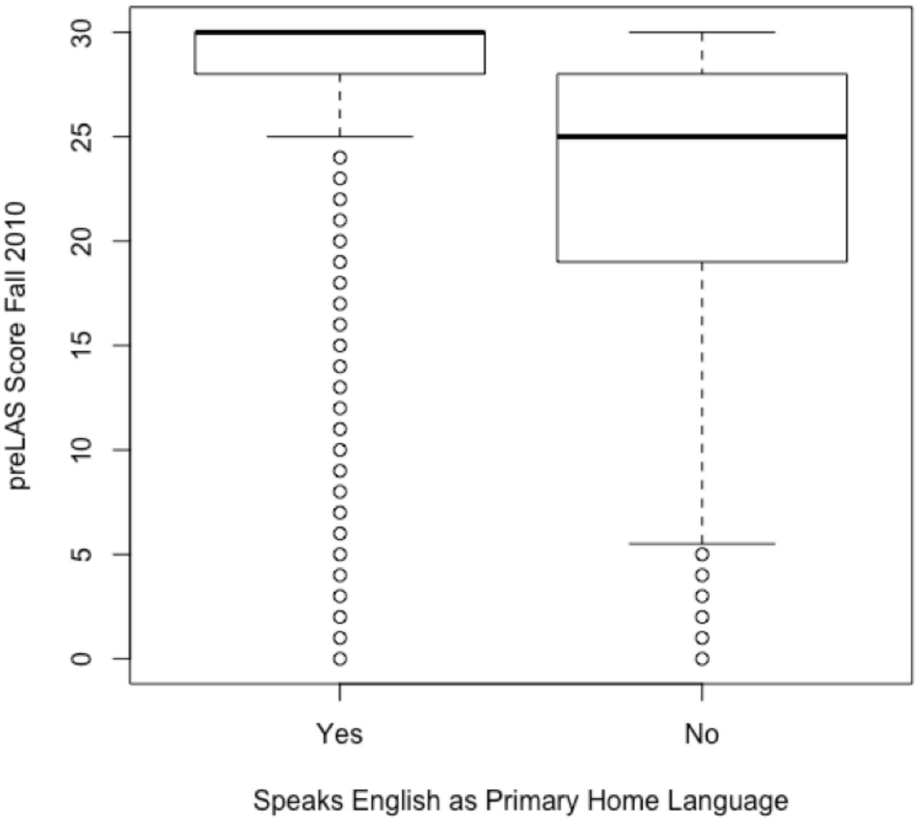

Figure 3. Weighted boxplots of Fall 2010 kindergarten sample *pre*LAS scores by primary home language.

Note. Parents of students who reported more than one primary home language were included in "Yes" for the purposes of this study. Figure produced using the *survey* package (Lumley, 2004; 2018).

Fall kindergarten teacher-rated attentional focus. Students' attentional focus was assessed using the attentional focus scale from the short form of the Children's Behavior Questionnaire (Putnam & Robertson, 2006). Teachers were asked to rate student behavior in response to scenarios over the previous six months. Ratings were provided on a seven-point scale (1 = *Extremely Untrue*, 7 = *Extremely True*; Tourangeau et al., 2015a). Higher scores represent greater attentional focus. Composite scores were prorated such that respondents needed four or more responses to have a scale score, which was calculated as the mean of the items. The scale has a reported reliability of .87 in the ECLS-K: 2011 data (Tourangeau et al., 2015a).

Fall kindergarten teacher-rated inhibitory control. Students' inhibitory control was assessed using the inhibitory control scale from the short form of the Children's Behavior Questionnaire (Putnam & Robertson, 2006). Teachers were asked to rate student behavior in response to scenarios over previous six months; ratings were provided on a seven-point scale of (1 = *Extremely Untrue*, 7 = *Extremely True*; Tourangeau et al., 2015a). Higher scores indicate higher levels of the ability to restrain behaviors. Composite scores were prorated such that respondents needed four or more responses to have a scale score, which was calculated as the mean of the items. The scale has a reported reliability of .87 in the ECLS-K: 2011 data (Tourangeau et al., 2015a).

Fall kindergarten cognitive flexibility. Cognitive flexibility was assessed using the Dimensional Change Card Sort (DCCS) task (Zelazo, 2006). This task measures a child's ability to shift problem solving in response to changing constraints or rules for the sorting task. For Fall kindergarten, the ECLS-K researchers administered the traditional table-top version, rather than the computerized National Institutes of Health (NIH) Toolbox version that was administered after first grade (Tourangeau et al., 2018). The assessment was comprised of one baseline task and then two tasks consisting of two different sorting rules. The raw number correct out of all trials was used in the current analyses. Given the relationship of cognitive flexibility to academic as well as other executive functioning outcomes, it is important to control for it to better isolate the relationship of working memory to mathematics. Reliability estimates for the table-top version of the DCCS are not reported; however, the equivalent NIH Toolbox computerized version for ages 3 to 15 has test-retest reliability of .92 (Weintraub et al., 2013).

Spring kindergarten working memory. Working memory was assessed using the Numbers Reversed subtest from the Woodcock-Johnson III Test of Cognitive Abilities (Woodcock, McGrew, & Mather 2001). In this task, the child listened to a series of numbers and was subsequently asked to repeat the series of numbers backwards to the assessor. The length of the digit span increased in blocks (i.e., multiple items with three digits, then the next block consistent of four-digit series, etc; Tourangeau et al., 2015a). The task was discontinued when the child reached the ceiling criterion. I used age-based standard scores in all analyses, which are designed to have a mean of 100 and standard deviation of 15 in the norming sample (Tourangeau et al., 2015a). The median reliability across ages in the norming sample is .87 (Schrank, McGrew, & Woodcock, 2001).

The Spring of kindergarten measurement was chosen for analysis predominantly for practical reasons, as a significant portion of students did not have valid working memory scores in the Fall of kindergarten. Students aged 62 months and younger who obtained a raw score of 0 or 1 did not obtain a standard score (Tourangeau et al., 2015a), and this would have significantly limited the sample size. Additionally, the first measurement in the growth modeling process (described more below) is the Spring of kindergarten; thus, utilizing the Spring measurement of working memory, though it may not address questions related skills at kindergarten entry, presents an accurate measure of the sustainability of the relationship between working memory and mathematics skills over time.

Procedure

Data from the current study were publicly available through the United States Department of Education's National Center for Educational Statistics website (NCES). Data were downloaded from the website as an ASCII file, which was then converted into a readable data file with Stata 15 syntax files NCES provides on its website. Data were partially cleaned in Stata then exported to R (R Core Team, 2018) for further manipulation and multiple imputation (more detail on this is provided below). After data manipulation and imputation in R, data were exported to Mplus (Múthen & Múthen, 2017) for further analysis. Because use of publicly available data from NCES in this format does not constitute research with human subjects, all research procedures were exempt from institutional review board (IRB) review by default.

The ECLS-K: 2011 utilizes a three-stage complex sampling design (Tourangeau et al., 2015a). In stage one, 90 primary sampling units (PSUs) consisting of counties were identified throughout the United States. Stage two consisted of identifying public and private schools that had kindergarten programs (or equivalent ungraded programs) for five-year-old children. Stage three consisted of sampling students within each school identified in stage two. Asian, Native Hawaiian, and Pacific Islander children were oversampled as a whole group (not as individual demographic groups; Tourangeau et al., 2015a). Due to this complex sampling design, traditional analyses that assume simple random sampling do not produce generalizable and unbiased estimates. As such, I completed all analyses using the appropriate weighting and design elements. More information on the weighting and estimation methods in this investigation is provided below in the Analysis section.

For Spring assessment data from Spring of first grade through Spring of fourth grade, sampling procedures were slightly adjusted. The sampling and recruitment procedures described above detail the methods used for the kindergarten base-year sample. These students were eligible for sampling in first through fourth grade Spring assessment waves. Again, I used appropriate sample weighting procedures in data cleaning and analysis to account for the design of sampling the longitudinal data, including non-response and students not eligible for direct assessment. More information on design and methodological limitations of sampling for Spring assessment data in grades first through four can be found in chapter four of the ECLS-K: 2011 Kindergarten – Fourth Grade User Manual (Tourangeau et al., 2018).

Missing data. Missingness may arise from a variety of sources. Enders (2010) describes three mechanisms of missingness, drawing on Rubin's (1976) work. I briefly summarize each of these three mechanisms below. Potentially the least plausible (yet most convenient) assumption of missingness is missing completely at-random (MCAR), in which missingness is not attributable to any observable characteristics about respondents and reflects truly random processes in nonresponse (e.g., a computer fluke prevents a response from being recorded). More realistic and still easy to handle with modern techniques is the missing at random (MAR) assumption, in which missingness on a response variable is associated with or attributable to other observed, measured characteristics. For example, families may be less likely to show up for a child therapy session and provide data due to factors related to socioeconomic status or demographics. The last mechanism, and the most difficult to handle, is the missing not at random assumption (MNAR). This may arise when missingness occurs because of the values respondents would report on the variable. A frequent occurrence of this in school settings (and the ECLS-K: 2011) is when schools do not administer English-based assessments to ELLs. Although one may partially account for the missingness by imputing assessment scores based on ELL status or ELP level, the primary reason the variable is missing is because it is highly likely an ELL with low ELP would score significantly lower on an English-based assessment than an ELL with higher ELP or an EPS. Under the MAR assumption, one may get closer to modeling MCAR data provided that an appropriate model of missingness is specified.

Multiple imputation offers a flexible, rigorous method to model missingness patterns (Rubin, 1987). The plausibility of appropriately accounting for the MCAR or MAR assumption in imputation also rests in part on the amount of missing data. Larger amounts of missing data induce more variability and uncertainty, making it harder to generate the distribution of data that would have been obtained with complete data from all respondents. Even in light of substantial portions of missing data, the major advantage to multiple imputation is the explicitly modeling of uncertainty in missingness in both the variance and estimates of missing values and the subsequent parameter estimates that are pooled across imputed datasets. In the current data, the sampling weights accounted for the majority of missing data due to nonresponse and attrition. However, missing data remained after weighting primarily for the teacher-report CBQ scales. This is likely primarily due to the item non-response rather than unit non-response (i.e., teachers did not complete enough items to obtain a total score though they partially completed the rating scales).

I used multiple imputation with predictive mean matching (PMM; Rubin, 1987) in the R package *mice* (van Buuren & Groothuis-Oudshoorn, 2011) to handle missing data. For the cognitive and academic measures there were missing very few data (i.e., mathematics, reading, and working memory). The convergence of the PMM model was very poor for these variables, so I imputed the mean of each group for each of these variables prior to conducting multiple imputation for the remaining missing data. All of the variables used in the current analyses were included in the imputation procedure. Auxiliary variables were also included to improve the accuracy of the imputation process (Enders, 2010). All missing values on independent and dependent variables were imputed, and I generated 20 imputed data sets each with 5 iterations. Table 3 provides information on each of the variables that were used in the imputation procedure, which I used primarily to account for missingness on the CBQ Attentional Focus and Inhibitory Control measures in the Fall of kindergarten. Some students were retained in kindergarten, and as a result those students were assigned a different variable for their CBQ measure in spring of first grade. Some adjustments to the teacher-level questionnaire had to be made for those students' measures to better reflect the kindergarten environment (Tourangeau et al., 2015b). As such, the spring of first grade CBQ Attentional Focus measure variable is titled X4ATTNFS in the data file and the same variable for retained students is titled X4KATTNFS. These two variables were collapsed into the same variable to create single spring of first grade CBQ variables to be used as an auxiliary variable in imputation (Enders, 2010). Students' retention status was not reflected in the cognitive assessment data (Tourangeau et al., 2018). For direct cognitive measures, students' grade corresponded to the year of the study. If a student was retained in kindergarten, they were still included in the grade one data. This pattern continued through the rest of the study such that students' grade reflected their year in the study, not the actual grade level in which they were enrolled. Figure 4 shows the density plots for each imputation of the CBQ scale scores. The correspondence between the red and blue lines indicates adequate performance of the PMM procedure in uncovering the sampling distribution of the scale scores.

Table 3
Variables Used in Multiple Imputation Model of Missing Data

Variable	Period of Measurement
Mathematics IRT Score	Fall/Spring of kindergarten and Spring grades 1-4
Reading IRT Score	Spring of kindergarten and grades 1-4
English Basic Reading Skills	Fall/Spring of kindergarten
*pre*LAS Simon Says Raw Score	Fall/Spring of kindergarten
*pre*LAS Art Show Raw Score	Fall/Spring of kindergarten
CBQ Attentional Focus	Fall/Spring of kindergarten and Spring grade 1
TMCQ Attentional Focus	Spring grades 2-4
CBQ Inhibitory Control	Fall/Spring of kindergarten and Spring grade 1
TMCQ Inhibitory Control	Spring grades 2-4
Dimensional Change Card Sort	Fall/Spring of kindergarten, Spring grade 1
WJ-III Numbers-Reversed	Spring of kindergarten and Spring grades 1-4
Age at Kindergarten Entry	Kindergarten
Socioeconomic Status (continuous)	Kindergarten
Primary Home Language (k – 1 dummy codes; Not English as reference group)	Kindergarten
Race/Ethnicity (k – 1 dummy codes; Hispanic as reference group)	Kindergarten
ELL Status	Fall of kindergarten
Full-Day Kindergarten	Fall/Spring of kindergarten
Parental Pre-K care	Kindergarten
Not First-Time Kindergartener	Fall of kindergarten
Assessed in Spanish	Fall/Spring of kindergarten, Spring grade 1
Primary Sampling Unit (k – 1 fixed effects)	One value for all study waves
Sampling Stratum (k -1 fixed effects)	“ ”
Sampling Weight	“ ”

Note. CBQ = Children’s Behavior Questionnaire; TMCQ = The Middle Childhood Questionnaire; WJ = Woodcock-Johnson; IRT = Item Response Theory

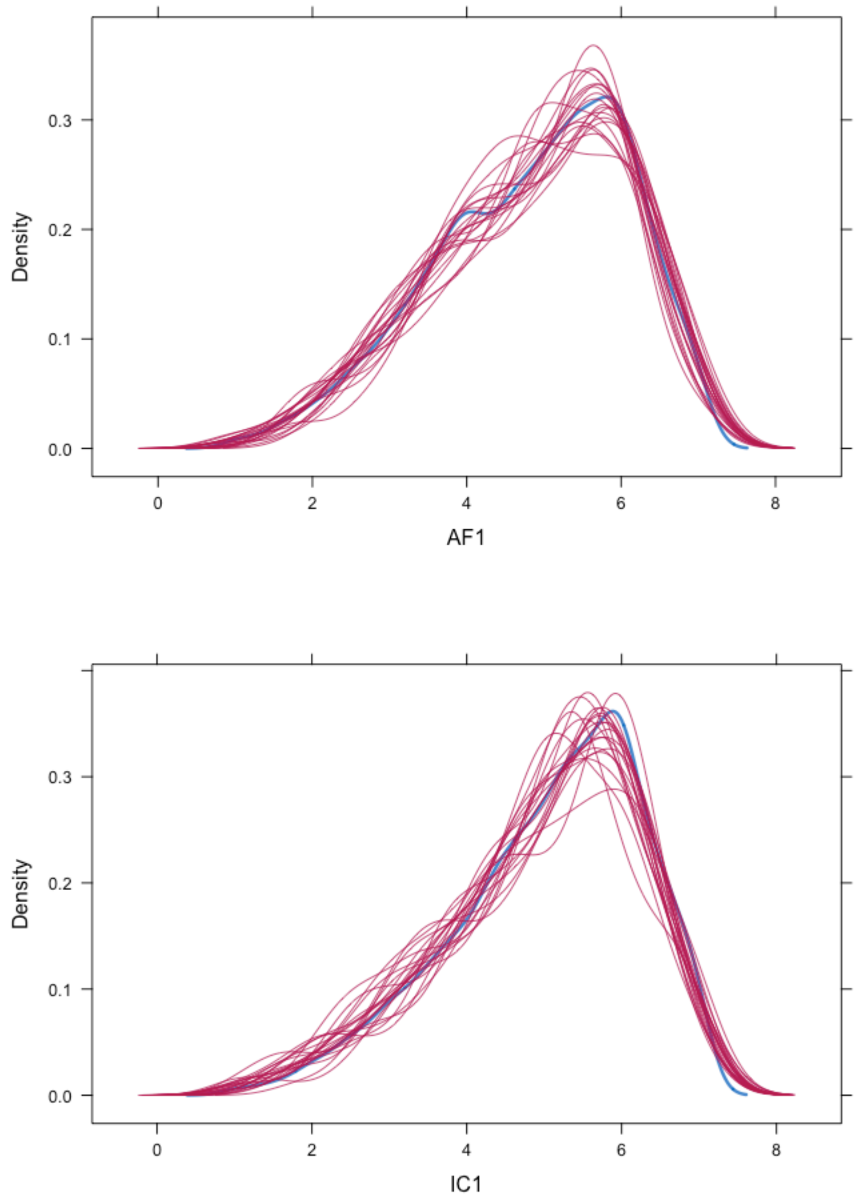

Figure 4. Density plots of 20 distributions of PMM-imputed values (red) and the observed sample distribution (blue) for the CBQ Attentional Focus (top) and Inhibitory Control (bottom) scale scores in the fall of kindergarten.

After removing students from the data who did not have the appropriate clustering variables (primary sampling unit and stratum) and had a sampling weight value of 0, there were few missing data. No more than approximately six percent of data were missing in the entire sample or in each of the EPS and ELL subgroups. Nevertheless, missingness that was not accounted for by the non-response weighting of the survey design may include other forms of bias that would not be appropriate to address through listwise deletion, which was present primarily in teacher-reported rating scales.

A few things are important to note about the imputation process. First, Fall 2010 *pre*LAS scores were not imputed. This was due to the fact that this variable was used in defining the ELLs of interest in the current study. Imputing a variable and creating a subset of data based on imputed values may misconstrue aspects of the population of interest by assigning people to categories to which they would not have belonged had they provided data, leading to potentially erroneous assumptions about students' demographic and linguistic characteristics. Missing values for primary home language were also removed prior to imputation. Although deleting these data is a significant limitation, it was more of a concern to accurately identify the student groups of interest than to increase statistical power, given the relatively large samples already available. Moreover, the use of sampling weights and survey design characteristics in the analyses excluded the majority of students with missing *pre*LAS scores and parent-reported home language from Fall 2010 (or Spring 2011 for home language). There were few (n = 18) students who were excluded because of missing *pre*LAS or home language reports but had valid sampling weight values. However, given the large amount of data, this is a small amount of sample loss. Other demographic and ecological variables were imputed, such as socioeconomic status, first-time kindergartener status, pre-kindergarten care arrangements, and kindergarten type (full/half day); however, because these were used as covariates and not for the purpose of creating subgroups, the imputed values were only useful in the context of controlling for these demographic and ecological characteristics rather than defining subgroups. The actual values of the control variables were not of substantive interest and were included to control confounds present in the kindergarten year that potentially relate to the level and course of mathematics performance and growth and its relationship with working memory. Third, many measures used in these analyses required students to obtain at least a minimum score to have a scorable assessment and speak English or Spanish. For cognitive assessment areas, the ECLS-K:2011 dataset includes flag variables that indicate whether a valid assessment score was present or not (e.g., the student attempted the assessment but did not obtain a minimum score or the student was ineligible due to speaking a language other than English or Spanish and did not pass the *pre*LAS screener). Students who did not provide a scoreable assessment based on the flag indicator in mathematics and reading at any time point, WJ-III Numbers Reversed assessments from the spring of kindergarten and later, and Dimensional Change Card Sort in fall and spring of kindergarten and spring of first grade were deleted. Deleting data from students who did not obtain minimum scores or were otherwise ineligible for assessment scores did not have a large impact on the sample size used in the current study; 1.2% (n = 63) students did not have valid assessment scores.

The last important aspect of the imputation is how I dealt with missingness on individual items of the CBQ. Prior research has indicated that imputing individual survey items then creating subscales based on the imputed values (rather than imputing subscale averages) provides more statistical power, though either approach induces similar levels of bias into the analytic model (Gotschall, West, & Enders, 2012). Given the relatively small proportion of missingness (approximately 6%), I imputed the preexisting scale score values of the CBQ measures provided in the ECLS-K: 2011 rather than individual items. Imputing scale scores also produced a more parsimonious imputation model in light of the number of other variables included in the model (i.e., *k* -1 dummy codes for strata and PSUs, race/ethnicity, and primary home language). Nevertheless, as a sensitivity analysis of the scale score imputation model, I also conducted an item-level imputation model using the item-level CBQ data for kindergarten and Spring of first grade (Tourangeau et al., 2018). Only these rounds of data were used to simplify the imputation model, given that adding all rounds of data for the CBQ and TMCQ attentional focus and inhibitory control items would require 78 additional variables (six items on each scale in kindergarten and first grade, and seven items on each scale in second through fourth grade). With the similarity in items, this may have caused substantial multicollinearity and convergence problems. Imputation results were pooled and analyzed with complex survey design elements in *survey* package in *R* (Lumley, 2004; 2018). Table 8 shows the estimates from the scale-level and item-level imputation procedures. Almost no differences were observed for the mean estimates. One aspect to note about the item-level imputation is the fact that the CBQ has the response option of "Not Applicable." In the prorated scale scores already provided in the dataset, these responses are treated as missing. There is likely a significant difference in why teachers would respond "Not Applicable" versus skipping the question altogether; "Not Applicable" likely reflects a systematic response (MNAR) whereas skipping an item could be truly random (MCAR), or at least predictable based on other variables (MAR). In the item-level imputation, these "Not Applicable" responses were imputed, but this could be problematic because it may imply responses were imputed that should not have existed in the first place (e.g., if the CBQ question was truly not applicable to the child). Thus, the CBQ scale score imputation model is advantageous because the available scores do not include “Not Applicable” options, so the responses reflect levels of attentional focus and inhibitory control that are based on relevant questions about the child’s behavior. Although imputing the prorated scale score sacrifices significant information about individual items, the prorated scores arguably reflect a more accurate picture of a child’s behavior among those who responded to enough questions than imputing responses that may truly have not been applicable to the child.

Table 4

Mean and Standard Deviation (SD) Estimates of CBQ Scale Scores Across Item- and Scale-Level Imputation

Scale	Item-Level Imputation		Scale-Level Imputation	
	M	*SD*	*M*	*SD*
ELL Attentional Focus	4.410	1.270	4.389	1.317
EPS Attentional Focus	4.796	1.249	4.793	1.273
ELL Inhibitory Control	4.649	1.232	4.729	1.312
EPS Inhibitory Control	4.933	1.190	4.994	1.244

Following the data cleaning and imputation procedures, the combined datasets were then split into 20 individual data files and read into M*plus* (version 8.1; Muthen & Muthen, 2017) for analysis. When conducting analyses on multiply imputed data, parameter estimates, standard errors, log-likelihood values, chi-square values, and the resulting fit indices are estimated across each of the imputed datasets and pooled into average estimates in a final model.

Analysis. I used a multiple-group latent change score modeling (MG-LCSM) approach to decomposing group differences in mathematics development. LCSMs offer a more nuanced approach to modeling differences over time than typical growth models, though LCSMs are nested with the larger growth modeling framework. In a LCSM, there are multiple forms of change: constant, proportional, and dual change (McArdle & Nesselroade, 2014; Petscher, Koon, & Herrara, 2016). Constant change is similar to the estimate provided in a simple linear growth model, though the interpretation of this parameter can become complicated in certain circumstances (discussed more below). Proportional change is the amount of change that occurs between individual waves of measurement relative to the level of the variable in the prior wave over and beyond what is accounted for by constant change. The constant change and proportional change models can be implemented on their own. Constant change constructed from change scores yields a slope parameter identical to that of the traditional linear growth curve. Proportional change on its own, while not frequently used in the literature, provides information about wave-to-wave change without assuming an underlying additive change trajectory that occurs across all waves.

Combining together constant and proportional change in a dual-change model (McArdle & Nesselroade, 2014; Petscher et al., 2016), the LCSM then captures the underlying constant (i.e., linear) component of growth while accounting for wave-to-wave-specific deviations from the linear form of change that are predicted based on one's prior level on the dependent variable (i.e., proportional change). The combination of these elements of change yields projected growth trajectories that capture nonlinear longitudinal relationships and provide greater detail as to when (and among what levels of performance) change is expected to occur.

Moreover, the proportional change components can be constrained to equality or estimated over time and across groups. Although fixing these proportional change components over time still yields non-linear growth trends, the implication is that one's prior level predicts subsequent growth to the same degree (measured as a proportion) across time. For example, it would suggest that students who have higher prior levels always change more in the subsequent time intervals, and the size of this effect (i.e., the amount of predicted change proportional to prior level) is equal across time. Freely estimating the proportional patterns over time, on the other hand, allows one to model non-constant autoregressive patterns in development. Petscher et al. (2016) demonstrated that modeling non-constant autoregressive patterns can be highly informative in the context of tracking academic growth within the RTI context, as predictive capability of academic measures progress monitoring and detecting adequate academic growth is crucial to providing adequate measurement of instruction and intervention response. Detecting varying individual differences in predicting change at different points thus helps detect variable patterns of development while accounting for the constant change that occurs across time. Although it is not expected that LCSMs will produce significantly different results in terms of the shape of change over time given prior work in this area (Grimm, Ram, & Estabrook, 2010; Roberts & Bryant, 2011), LCSMs provide more insight into the relationships between students' static performance and the extent of their growth across time while simultaneously accounting for the general pattern of additive change over the developmental period (kindergarten through fourth grade).

As a structural equation model, LCSMs can be fit as a multiple group model (MG-LCSM) and parameters can be fixed or estimated across groups depending on hypotheses about the values of structural parameters. A multiple-group model then allows the testing of invariance of these parameters across time and groups. A variety of techniques have been used to study the differential longitudinal patterns of development including multiple-group growth models; however, the use of the MG-LCSM facilitates the detection of differential patterns in non-constant autoproportional change and a method of explicitly testing the invariance of change parameters across groups.

Therefore, the goal of Study 1 is to test invariance of change parameters between ELLs and EPSs in a MG-LCSM. In particular, the invariance of dual-change parameters (i.e., constant and proportional change) across ELLs and EPSs in addition to covariate relationships in a conditional MG-LCSM will be tested. Multiple models were considered in testing hypotheses of change, both unconditional and conditional (i.e., controlling for a number of baseline covariates). As aforementioned, ELLs were identified as those students who spoke a language other than English at home and were low in English proficiency in the Fall of kindergarten (i.e., scored lower than 23 on the Fall 2010 administration of the *pre*LAS). Although this definition of ELL using the ECLS-K:2011 differs from prior studies using the 2011 (e.g., Greenfader, 2017; Hartano et al., 2018; Wang, 2017; Wilkinson, 2017) and the 1999 ECLS-K cohort (e.g., Halle et al., 2012; Roberts & Bryant, 2011), the main goal of this study is to detect differences in change patterns among students who were not initially proficient in English while making a conscientious effort to define the subpopulation of initially low ELP ELLs based on *pre*LAS proficiency relative to peers who speak English, at least in part, as a primary home language. Although change patterns among initially English-proficient second-language learners are of major interest, these patterns have in part been documented in the literature (e.g., Halle et al., 2012; Roberts & Bryant, 2011).

Data inspection. As with basic regression analyses, violations to assumptions of structural equation models can significantly impact standard error efficiency and bias parameter estimates. The basic assumption of SEM is multivariate normality. I assessed this using the *MVN* package in R (Korkmaz, Goksuluk, & Zararsiz, 2014). Results from this procedure indicated significant violations to multivariate normality when examining only the dependent variables (i.e., mathematics IRT scores from kindergarten through grade four) as well as when including the entire set of independent and dependent variables. The heavily-tailed distributions of the mathematics IRT scores is evident in the descriptive boxplots presented previously. This skew and kurtosis of the IRT distributions is a natural product of the scale's construction (Najarian et al., 2018a, 2018b). As such, all analyses employed robust estimation techniques to account for violations to normality. Another factor that can impact SEM results is the presence of significant outliers, both univariate and multivariate. I assessed univariate outliers of growth curves through inspection profile plots of each language group sample. Although there was significant variability in wave-specific distributions, few individual raw profile plots appear to be significant outliers such that they could be problematic for model estimation (in terms of fitting growth curves to individual data points). However, at wave five (fourth grade) there were five students (one ELL, four EPSs) who scored > 5 *SD* below the mean of 109.29 (*SD* = 14.86). Because of these students' significant univariate distance from the mean at that specific time point, additional care was taken to inspect models after fitting to examine the impact, if any, these data points had on model estimation. Additionally, for the conditional models, one extreme univariate outlier was present for the Fall of kindergarten mathematics IRT score, which was used as a covariate. One EPS obtained a score of 139.16, which was nine *SD*s above the Fall kindergarten sample mean of 35.15. The extremity of this data point warranted further investigation after model fitting to determine its leverage in the model.

Complex survey elements. The complex design of the survey necessitated weighted analyses to produce generalizable population estimates and standard errors corrected for heteroskedasticity due to clustering within PSUs and strata. I used the Taylor method of variance estimation to produce cluster-robust standard errors using the TYPE = COMPLEX command in M*plus* (Múthen & Múthen, 2017). By default, complex survey estimation in M*plus* employs robust maximum likelihood estimation (MLR), which uses a standard error correction procedure similar to the Yuan-Bentler "sandwich" estimator (Múthen & Múthen, 2017). This robust estimation also aids in mitigating violations of multivariate nonnormality in standard error estimation. In all analyses, I employed the sampling weight for Spring measurements between kindergarten and grade four along with the Fall of kindergarten (W8C18P_8T180) with the corresponding primary sampling unit (W8C18P_8T1PSU) and stratum (W8C18P_8T18STR).

Study 1 Results

Descriptive Statistics

Descriptive statistics on the independent and dependent variables in this study are presented in Tables 5 through 7. Additionally, Table 8 provides proportions of demographic characteristics for each group with corresponding tests of proportions for each group to assess baseline equivalence and differences for each characteristic. Table 9 shows t-tests for baseline differences in continuous covariates used in the MG-LCSM modeling.

Table 5
Overall Sample Descriptive Statistics

	M	*SD*	Min	Max	Proportion Missing
EBRS - Fall K	13.60	4.35	0.00	20.00	.000
IRT Mathematics - Fall K	35.15	11.52	10.25	139.16	.000
IRT Mathematics – Sp. K	49.12	12.39	13.58	98.29	.000
IRT Mathematics – Sp. Gr. 1	73.40	16.93	21.91	131.82	.001
IRT Mathematics – Sp. Gr. 2	90.16	15.99	14.79	143.97	.001
IRT Mathematics – Sp. Gr. 3	102.55	15.18	40.25	144.25	.002
IRT Mathematics – Sp. Gr. 4	109.79	14.86	25.22	139.06	.002
WJ - III Numbers Reversed – Spring K	95.77	16.61	41.00	157.00	.000
DCCS – Fall K	14.43	3.13	0.00	18.00	.000
CBQ Attentional Focus – Fall K	4.77	1.28	1.00	7.00	.035
CBQ Inhibitory Control – Fall K	4.98	1.25	1.00	7.00	.038
Age at Kindergarten Entry	66.35	4.48	39.83	86.87	.002
Socioeconomic Status – Fall/Sp. K	-0.14	0.76	-2.33	2.6	.002
*pre*LAS Combined Score - Fall K	27.78	4.63	0.00	30	.000
Unweighted *N*			5,014		
Weighted *N*			3,432,073		

Note. Missingness proportions are not weighted. EBRS = English basic reading skills, IRT = item response theory, WJ = Woodcock-Johnson, DCCS = Dimensional Change Card Sort, CBQ = Children's Behavior Questionnaire, K = Kindergarten.

Table 6

EPS Sample Descriptive Statistics

	M	*SD*	Min	Max	Proportion Missing
EBRS – Fall K	13.95	4.07	0.00	20.00	.000
IRT Mathematics - Fall K	35.88	11.34	10.25	139.16	.000
IRT Mathematics – Sp. K	49.82	12.17	13.58	98.29	.000
IRT Mathematics - Sp. Gr. 1	74.33	16.69	21.91	131.82	.001
IRT Mathematics - Sp. Gr. 2	90.93	15.68	14.79	143.97	.001
IRT Mathematics - Sp. Gr. 3	103.18	14.96	40.25	144.25	.002
IRT Mathematics - Sp. Gr. 4	110.36	14.69	25.22	139.06	.002
WJ - III Numbers Reversed – Spring K	96.44	16.40	41.00	157.00	.000
DCCS – Fall K	14.58	2.99	0.00	18.00	.000
CBQ Attentional Focus	4.79	1.27	1.00	7.00	.035
CBQ Inhibitory Control	5.00	1.24	1.00	7.00	.038
Age at Kindergarten Entry	66.44	4.45	39.83	86.87	.002
Socioeconomic Status	-0.09	0.75	-2.07	2.6	.002
*pre*LAS Combined Score - Fall K	28.69	2.32	5.00	30.00	.000
Unweighted *N*			4,711		
Weighted *N*			3,226,117		

Note. Missingness proportions are not weighted. EBRS = English basic reading skills, IRT = item response theory, WJ = Woodcock-Johnson, DCCS = Dimensional Change Card Sort, CBQ = Children's Behavior Questionnaire, K = Kindergarten.

Table 7

ELL Weighted Sample Descriptive Statistics

Variable	*M*	*SD*	Min	Max	Proportion Missing
EBRS – Fall K	8.05	4.63	0.00	20.00	.000
IRT Mathematics - Fall K	23.73	7.60	9.79	48.54	.000
IRT Mathematics – Sp. K	38.24	10.58	14.13	76.11	.000
IRT Mathematics - Sp. Gr. 1	58.93	13.88	25.88	103.40	.000
IRT Mathematics - Sp. Gr. 2	78.10	16.02	23.45	115.43	.000
IRT Mathematics - Sp. Gr. 3	92.66	15.14	43.95	123.92	.001
IRT Mathematics - Sp. Gr. 4	100.87	14.77	28.16	130.95	.001
WJ - III Numbers Reversed – Spring K	85.22	16.46	51.00	134.00	.000
DCCS – Fall K	12.09	4.12	0.00	18.00	.000
CBQ Attentional Focus – Fall K	4.39	1.32	1.00	7.00	.063
CBQ Inhibitory Control – Fall K	4.73	1.32	1.00	7.00	.056
Age at Kindergarten Entry	65.02	4.83	39.10	81.07	.003
Socioeconomic Status – Fall/Sp. K	-0.84	0.54	-2.33	1.61	.003
*pre*LAS Combined Score - Fall K	13.56	7.57	0.00	22.00	.000
Unweighted *N*			205,956		
Weighted *N*			303		

Note. Missingness proportions are not weighted. EBRS = English basic reading skills, IRT = item response theory, WJ = Woodcock-Johnson, DCCS = Dimensional Change Card Sort, CBQ = Children's Behavior Questionnaire, K = Kindergarten.

Table 8

Study One Weighted Demographic Proportions and χ^2 Difference Tests

Variable	EPS	ELL	χ^2 Sig.	Total	Proportion Missing	
					EPS	ELL
Primary Home Language: English	.88	--	--	.83	.000	.000
Primary Home Language: Not English	.11	1.00	***	.16	.000	.000
Primary Home Language: Cannot Choose	.01	--	--	.01	.000	.000
Black/African American	.14	.007	***	.13	.000	.000
White	.54	.006	***	.51	.000	.000
Hispanic, Race Specified	.21	.91	***	.25	.000	.000
Hispanic, No Race Specified	.003	.01	*	.004	.000	.000
Asian	.04	.05		.04	.000	.000
Native Hawaiian/Pacific Islander	.003	.01		.004	.000	.000
American Indian/Alaskan Native	.01	.002	*	.01	.000	.000
Two or More Races	.043	--	--	.04	.000	.000
Female	.49	.50		.49	.000	.000
Not First-Time Kindergartener	.05	.05		.05	.002	.003
Parental Pre-K Care	.20	.38	***	.21	.004	.003
Full-Day Kindergarten (Fall)	.80	.91	*	.81	.001	.007
Full-Day Kindergarten (Spring)	.81	.93	**	.82	.000	.000
Unweighted *N*	4,711	303		5,014		
Weighted *N*	3,226,117	205,956		3,432,073		

Note. Missingness proportions not weighted. Primary home language measured in kindergarten. *$p < .05$, **$p < .01$, ***$p < .001$

Table 9
T-Tests of Baseline Mean Differences

Variable	EPS *M*	ELL *M*	M_{Diff} (*SE*)
EBRS – Fall K	13.95	8.05	5.90 (0.31)***
IRT Mathematics – Fall K	35.88	23.73	12.15 (0.72)***
WJ – III Numbers Reversed – Spring K	96.44	85.22	11.22 (1.05)***
DCCS– Fall K	14.58	12.09	2.49 (0.29)***
CBQ Attentional Focus – Fall K	4.79	4.39	0.40 (0.09)***
CBQ Inhibitory Control – Fall K	5.00	4.73	0.26 (0.07)***
Age at Kindergarten Entry	66.44	65.02	1.41 (0.45)**
Socioeconomic Status – Fall/Sp. K	-0.09	-0.84	0.75 (0.04)***

Note. All analyses weighted. EBRS = English basic reading skills, IRT = item response theory, WJ = Woodcock-Johnson, DCCS = Dimensional Change Card Sort, CBQ = Children's Behavior Questionnaire, K = Kindergarten, IRT = item response theory. ** $p < .01$, *** $p < .001$

Figure 5 shows profile plots of 30 randomly sampled ELLs and 30 randomly sampled EPSs (they were sampled separately to obtain exactly 30 students from each language group). Box plots for EPSs and ELLs across assessment waves are presented in Figures 6 and 7, respectively. Examining the box and profile plots descriptively, it is clear that the variability and mean level of change across assessment waves is markedly different between the two groups.

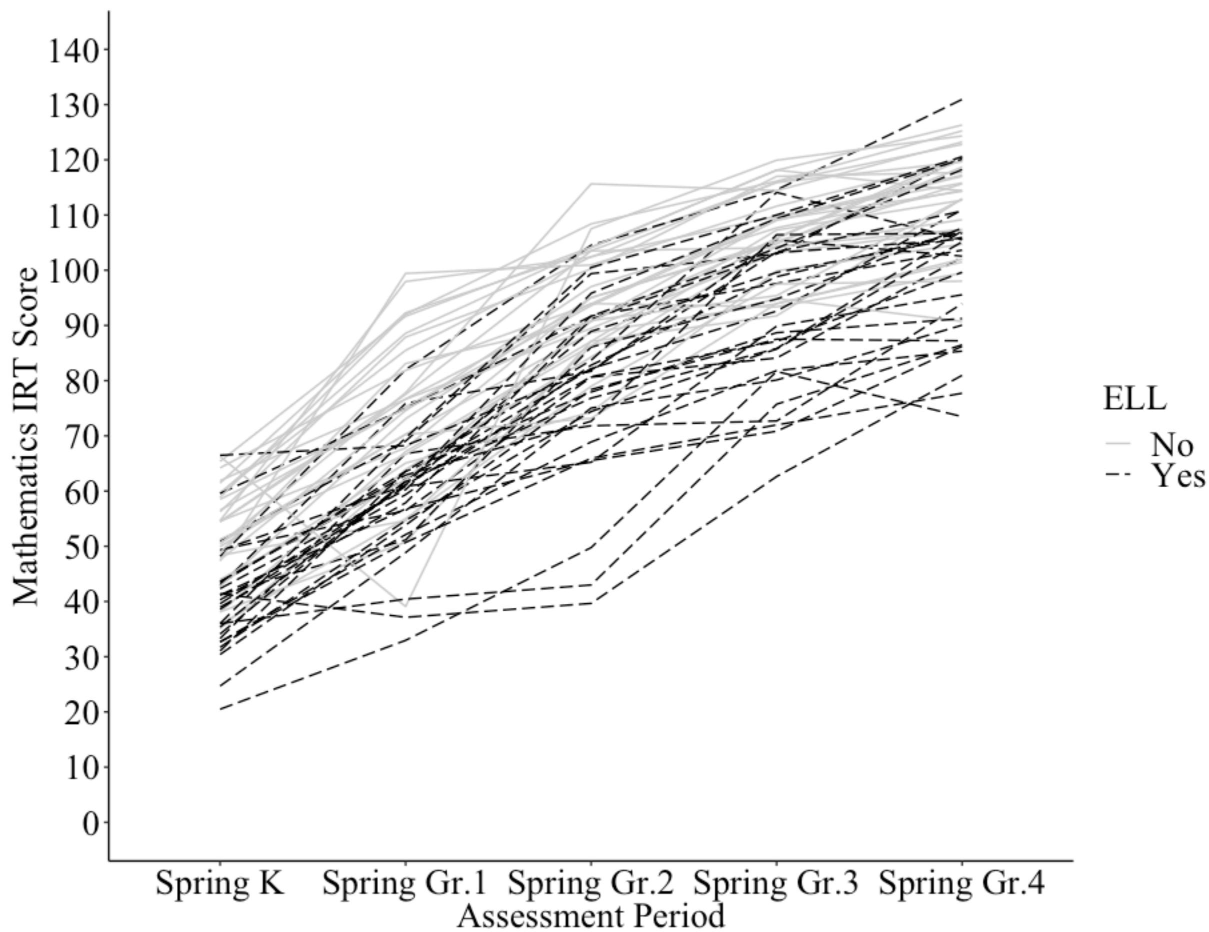

Figure 5. Profile plot of mathematics achievement between Spring of kindergarten and Spring of fourth grade for 30 randomly sampled EPSs and 30 randomly-sampled ELLs.

Note. Data randomly sampled separately within groups then combined to create plots. Figure produced in *ggplot2* (Wickham, 2016).

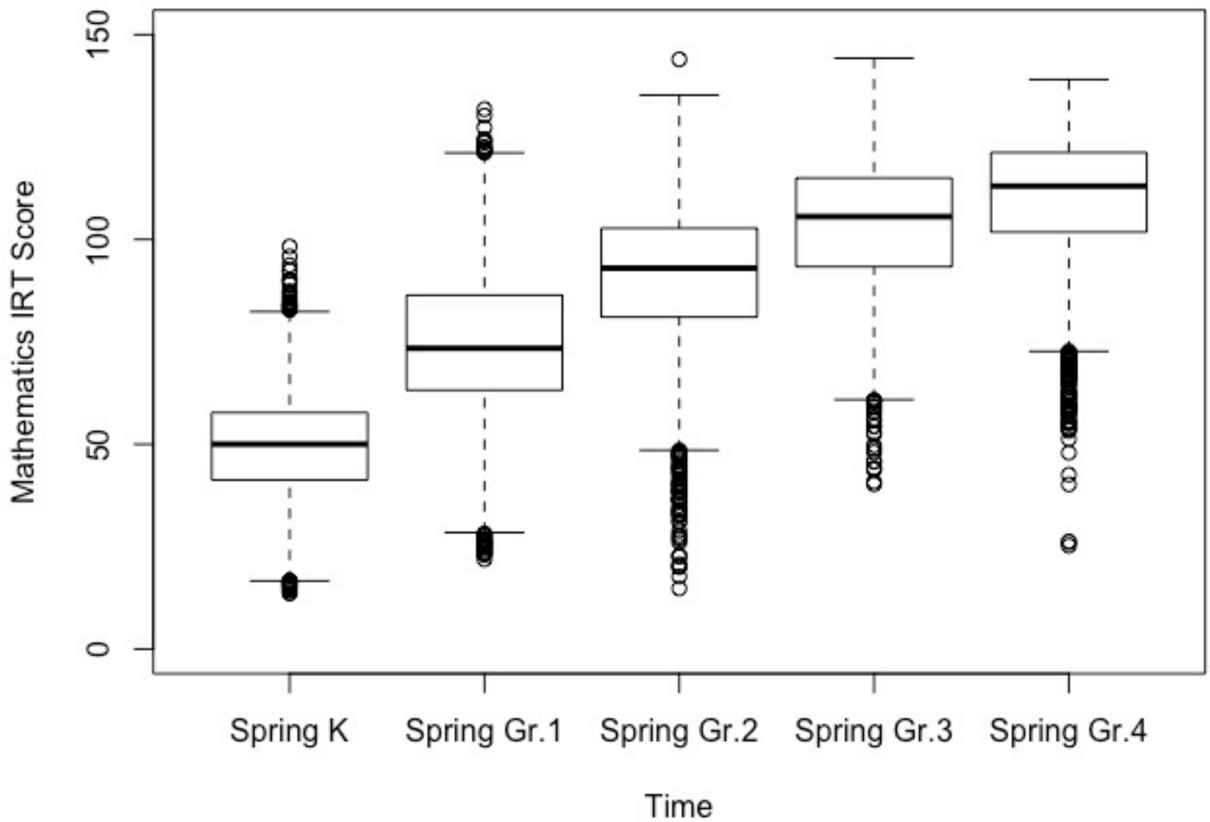

Figure 6. Weighted boxplot of EPS students' mathematics IRT scores by assessment wave.

Note. Figure produced using the *survey* package (Lumley, 2004; 2018).

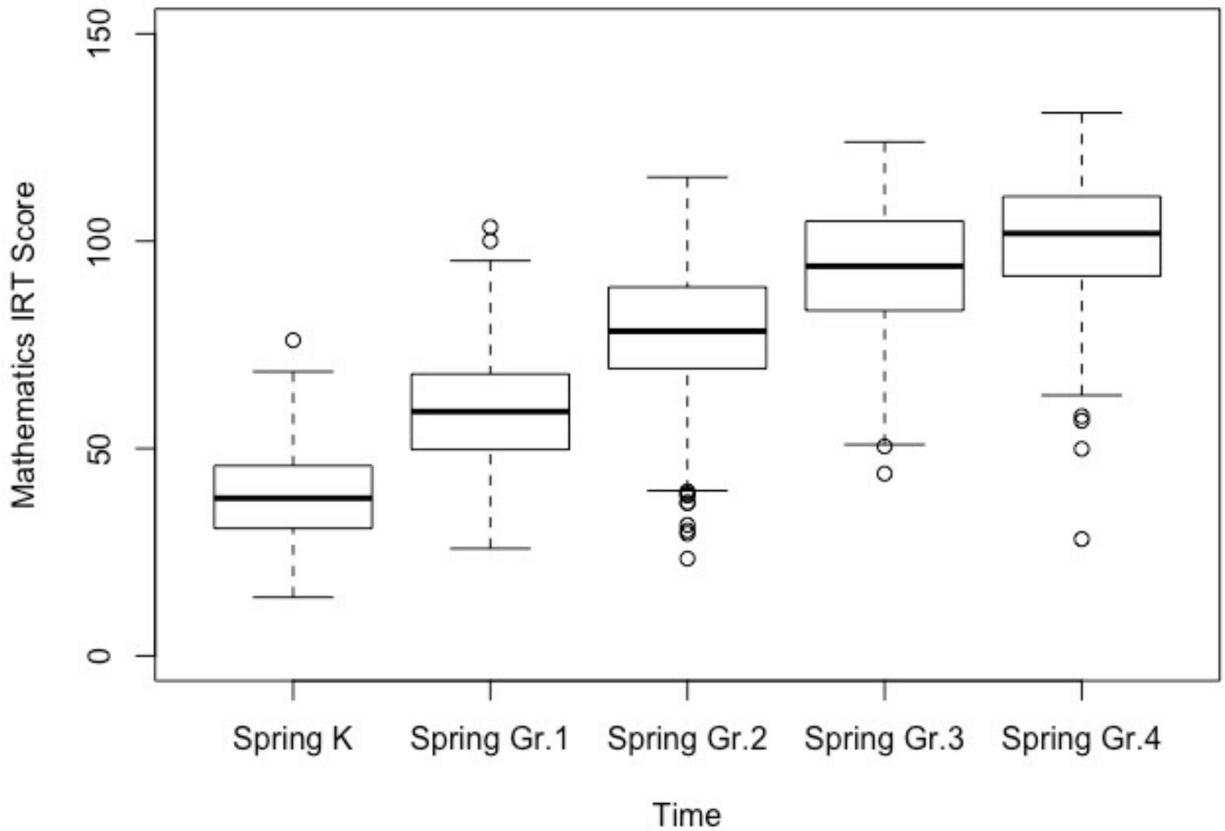

Figure 7. Weighted boxplots of ELL students' mathematics IRT scores by assessment wave.

Note. Figure produced using the *survey* package (Lumley, 2004; 2018).

Unconditional MG-LCSM

To begin to test the primary research question of the study, a series of unconditional MG-LCSMs were tested. I used the Bayesian Information Criterion (BIC) to assess model fit improvement. Starting with a single-group model, I then estimated a multiple-group model with all parameters constrained across groups. Following that, I iteratively and systematically freed parameters to test the improvement in fit. The model fitting results are presented in Table 10. Figure 8 is a path diagram of the final selected model based on the fit indices in Table 10.

Table 10
Unconditional MG-LCSM BIC Values

	BIC	ΔBIC from Model 1
Model 1		
Single group model (residuals freed over time, proportional change constrained over time)	179995.635	--
Model 2		
All change parameters constrained (residuals and proportional change constrained across time and groups)	180989.506	993.871
Model 3		
Intercept freed (residuals and proportional change constrained across time and groups)	180807.937	812.302
Model 4		
Intercept + proportional change + residuals freed across time and groups	**178922.785**	**-1072.85**
Model 5		
Intercept + slope + residuals + variances/covariances freed	178936.862	-1058.773
Model 6		
Intercept + slope + residuals + variances/ covariances + proportional change freed	178932.065	-1063.570

Note. Bolded values indicate the final model selected based on the lowest BIC value compared to Model 1.

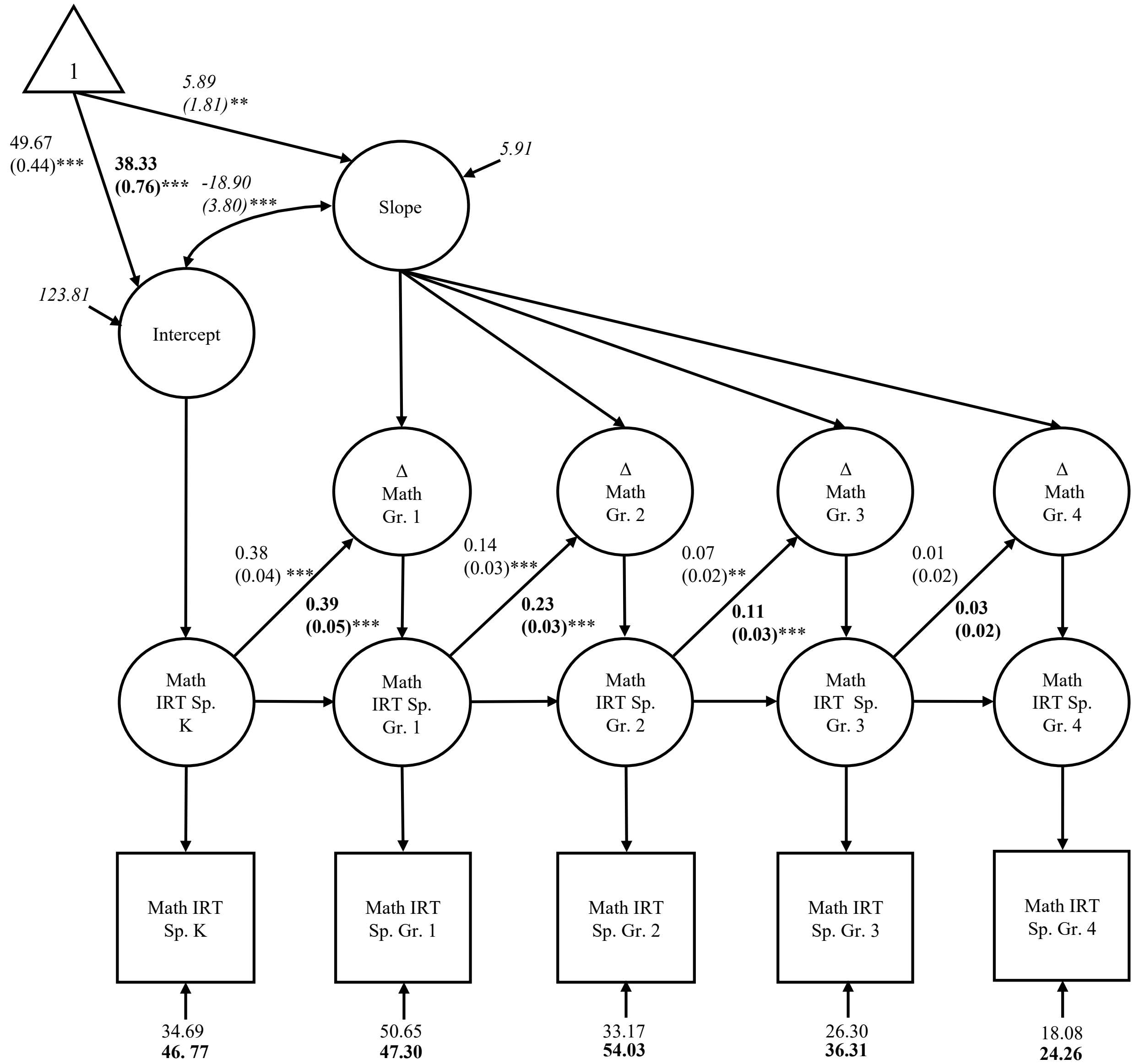

Figure 8. Multiple-group unconditional latent dual-change model of mathematics.

Note. Factor loadings for ELLs are bolded. Standard errors in parentheses. Italicized coefficients denote the parameter is fixed across groups. All unlabeled paths fixed to 1. **p* < .05, ***p* < .01, ****p* <.001

The final model selected based on Table 10 (χ^2 [df] = 363.487 [16]) generally evidenced adequate fit based on the root mean square error of approximation (RMSEA = 0.093), comparative fit index (CFI = 0.986), Tucker-Lewis index (TLI = 0.969), and standardized root mean squared residual (SRMR = 0.133).[2] Based on the parameters in Figure 8, ELLs and EPSs differed significantly in their intercept values, suggesting that in the Spring of kindergarten significant gaps in mathematics achievement were evident. This model indicated that constant change rates (slope) and intercept-slope variances and covariances were equal across groups, suggesting that the gaps observed in kindergarten generally persisted through the end of grade 4. In both groups, students who score higher tend to grow more, as indicated by the significant, positive proportional change values coupled with the positive constant change. However, this autoproportional effect tapers off by wave five, suggesting that individuals' mathematics IRT scores in grade three are not significantly predictive of how much they change through fourth grade. This model does not account for other factors that may explain group differences, though, so it is difficult to assess the extent to which the differences observed here are attributable to other confounders. As such, I considered a series of conditional models to address Research Question 1.

Conditional MG-LCSMs

Covariates were included into the MG-LCSM to analyze the extent to which the predictors of interest improved model fit and helped explain group differences. In all conditional models, the intercept, proportional change, and residual parameters were unconstrained across

[2] The typical independence model used to calculate indices such as the CFI and TLI is inappropriate for growth/latent change modeling (Widaman & Thompson, 2003). I calculated the CFI and TLI using a multiple-group intercept-only model with residuals constrained across group and time as the null model rather than the default independence model (χ^2 [df] = 24,223.455 [35]).

.

groups based on the findings from the unconditional MG-LCSM, though constant change and intercept-slope variances/covariances remained fixed across groups. Although it is likely that the inclusion of covariates would change the structural relationships across groups, the goal of analyzing the conditional model is to detect the extent to which covariates explain variation in growth and improve model fit directly compared to the baseline unconditional model. Continuous predictors were centered at the grand mean for the entire sample to aid in interpretation and detect deviations from the overall weighted sample mean values (except the SES variable since it was already standardized to have $M = 0$ and $SD = 1$ in the full sample).

First, a model with all intercept and slope regressions freed between groups was fit. The unconstrained conditional MG-LCSM showed better a BIC (172315.683) than the unconditional model. Figure 9 shows path diagram of the within-group conditional MG-LCSM. This type of model essentially assumes that ELL status moderates the relationship between mathematics growth and all of the covariates in the model. An alternative perspective is that, because this is a multiple group model that is estimating each group's model separately, this is equivalent to modeling two separate conditional LCSMs. Therefore, this model does not remove between-group differences, but rather provides a way to remove within-group differences to inspect the remaining between-group differences. Table 11 provides the regression parameter estimates and fit indices for the overall MG-LCSM. The aforementioned Fall of kindergarten mathematics score outlier did not affect model estimation, so I retained it in the analyses. Additionally, the five significant univariate outliers identified in wave five did not meaningfully alter regression or latent change estimates.

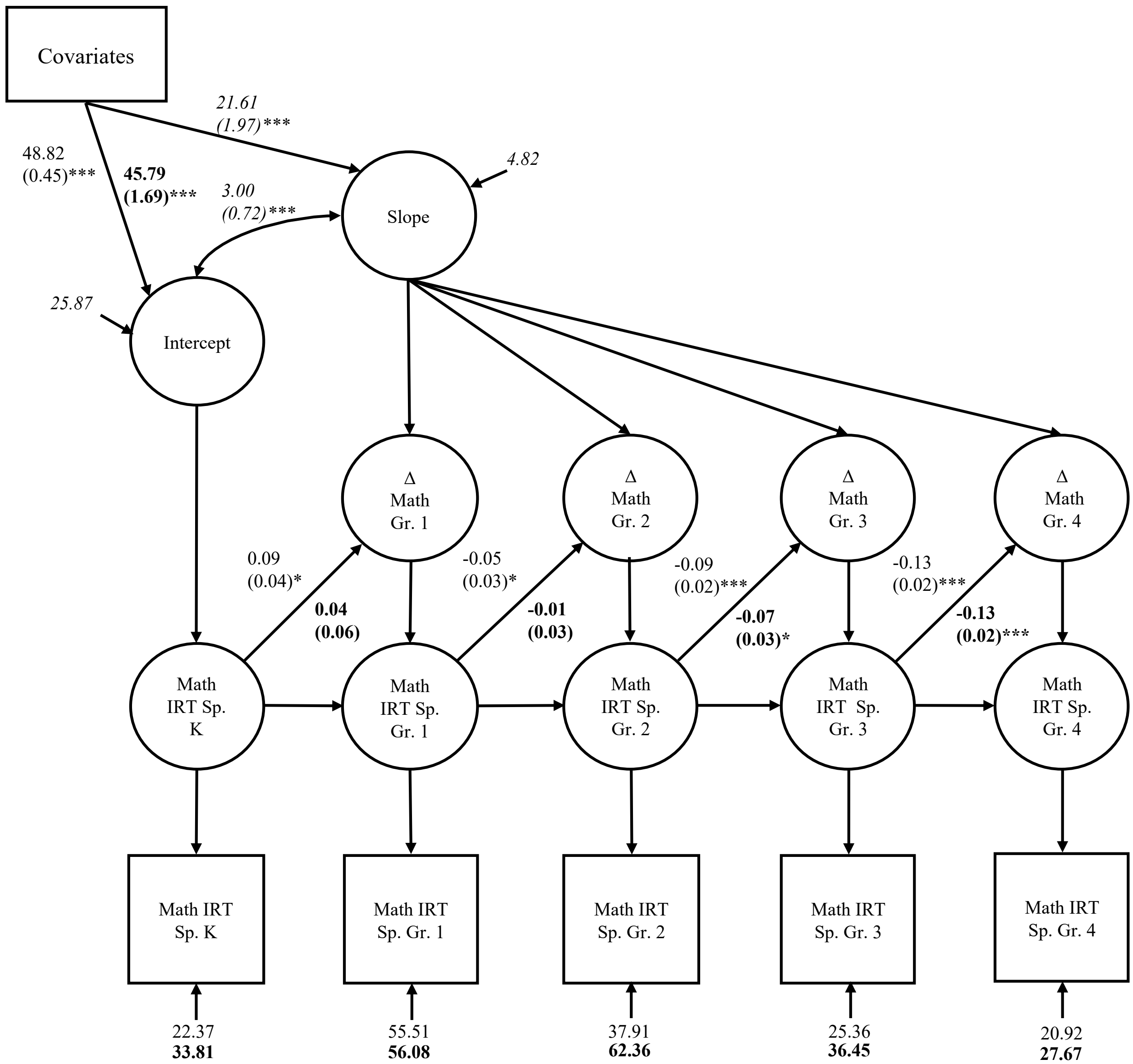

Figure 9. Conditional MG-LCSM with covariate regressions unconstrained across groups.

Note. ELL parameter estimates are displayed in bold. Standard errors in parentheses. Italicized coefficients are fixed across groups. All unlabeled paths fixed to 1. **p* < .05, ***p* < .01, ****p* <.001

Table 11

Multiple Group Conditional Latent Change Score Results - Regressions Freed Across Groups

	Regression Coefficients (*SE*)			
	ELL		EPS	
Predictor	Intercept	Slope	Intercept	Slope
(Intercept)	45.786 (1.699)***	21.639 (1.974)***	48.824 (0.469)***	21.639 (1.974)***
Sp. K Working Memory	0.167 (0.048)**	-0.003 (0.011)	0.158 (0.008)***	0.019 (0.007)**
Fall K DCCS	0.268 (0.107)*	0.115 (0.064)	0.177 (0.048)***	0.098 (0.029)**
Fall K Attentional Focus	1.059 (0.604)	0.579 (0.244)*	0.588 (0.158)***	0.209 (0.075)**
Fall K Inhibitory Control	0.070 (0.6)	-0.293 (0.252)	0.308 (0.137)*	-0.082 (0.067)
Fall K EBRS	0.202 (0.135)	0.052 (0.043)	0.080 (0.046)	-0.029 (0.018)
Fall K Full-Day	3.298 (1.354)*	-0.713 (0.669)	0.842 (0.438)	-0.551 (0.169)**
Parental Pre-K Care	0.295 (1.088)	-0.411 (0.385)	-0.236 (0.372)	0.163 (0.158)
Age of Kindergarten Entry	0.183 (0.173)	-0.146 (0.069)*	0.173 (0.039)***	-0.088 (0.016)***
Not First-Time Kindergartener	4.671 (2.694)	-1.650 (1.082)	0.900 (0.683)	-2.546 (0.428)***
Socioeconomic Status	0.474 (0.857)	-0.045 (0.402)	0.929 (0.198)***	0.481 (0.109)***
Non-Hispanic	-0.637 (1.657)	2.266 (0.815)**	0.129 (0.391)	-0.194 (0.175)
Female	-0.624 (1.066)	-1.046 (0.348)**	-1.001 (0.212)***	-1.233 (0.14)***
Fall K Mathematics IRT	0.541 (0.093)***	0.016 (0.04)	0.681 (0.018)***	-0.007 (0.017)
R^2	.68	.27	.81	.21
RMSEA	0.054			
SRMR	0.039			
BIC	172314.811			
χ^2 (*df*)	783.415 (94)			

Note. *$p < .05$, **$p < .01$, ***$p < .001$

After conditioning the MG-LCSM on the set of predictors and freely estimating the slope and intercept regressions within each group, the model-implied trajectories remained similar. EPSs had a higher intercept (nearly one-fifth *SD*s) in the Spring of kindergarten and showed somewhat different change patterns. EPSs continued to gain more throughout first grade and show slight deceleration in growth through fourth grade. ELLs show a mostly linear change pattern until grade two when the autoproportional effect increases and then nearly doubles again through fourth grade.

Though the last model conditions on the covariates, it does not control for between-group differences in the covariates and serves only as an intermediate model to examine conditional change patterns. A second conditional MG-LCSM model was estimated with covariate relationships constrained across groups, which controls for between-group differences in the covariate effects on the MG-LCSM intercepts and slopes. This allows detection of the extent to which change components differ between ELLs and EPSs after removing between-group differences in baseline characteristics. Figure 10 displays the MG-LCSM as a path model, and Table 12 displays the regression coefficients and fit indices. Based on the change coefficients in Figure 10, EPSs again continue to show consistent, significant autoproportional effects through fourth grade, the magnitude of which remain somewhat similar though the direction changes after first grade. ELLs show similar patterns after grade one; however, change remains relatively linear in earlier years (i.e., prior performance is not significantly predictive of change beyond the constant change in kindergarten or first grade). Controlling for these baseline covariates explains all of the ELL-EPS achievement gap at the end of kindergarten. That said, conditioning on these achievement, cognitive, and demographic characteristics does not significantly change group differences in change. Compared to the baseline unconditional model, however, this final conditional model has weaker autoproportional effects, indicating that the covariates potentially explain some of the patterns of intraindividual change. This may be due in part because of the significant difference in the intercept-slope covariance as well as the conditional slope (constant change) value (which is larger), suggesting that after conditioning on these covariates, a larger portion of change occurs through the conditional slope (constant change) value. Again, none of the significantly outlying data substantially impacted model estimation.

To visually compare the unconditional and two conditional models, Figure 11 shows model-implied trajectories for each of the models. As noted in the figure, Fall kindergarten mathematics is held at 0 (i.e., the grand mean) and all other covariates are held at each group's means. These trajectories show the estimated growth that occurs when each group enters kindergarten performing identically in mathematics but when they are held group-specific levels of all other covariates. Each trajectory is estimated at the mean of each dummy variable, which indicates the proportion of individuals in the data with a value of 1 on each dummy variable. This aids in extracting a true mean trajectory for each group rather than mean trajectories among specific groups (i.e., holding dummy variables at 1 or 0).

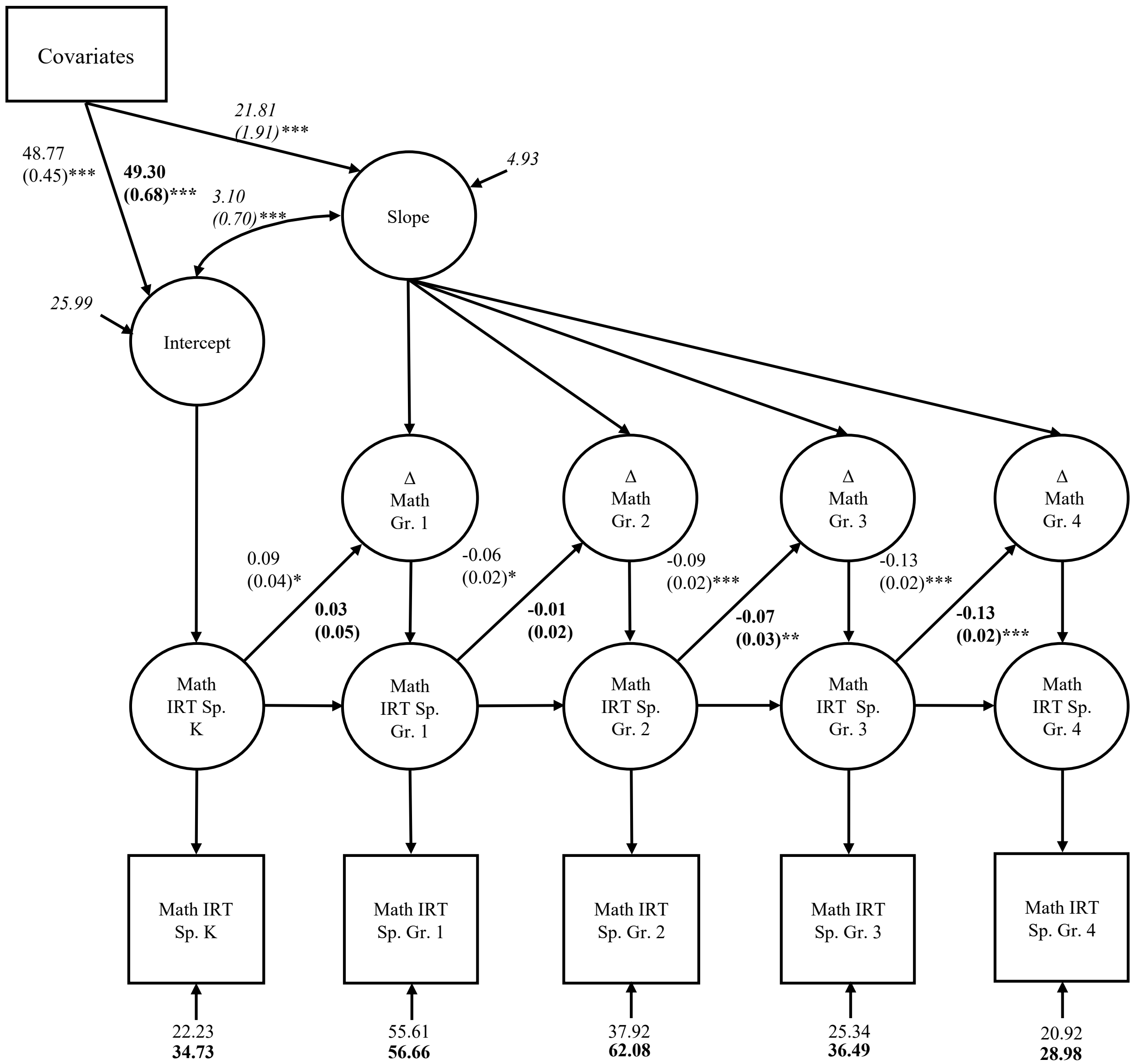

Figure 10. MG-LCSM estimates with regression parameters constrained to equality across groups.

Note. ELL parameter estimates are displayed in bold. Italicized coefficients denote it is fixed across groups. All unlabeled paths fixed to 1. **p* < .05, ***p* < .01, ****p* <.001

Table 12
Multiple Group Conditional Latent Change Score Results with Regressions Constrained Across Groups

	Regression Coefficient (SE)	
Predictor	Intercept	Slope
(Intercept)	48.765 (0.452)***/**49.295 (0.681)*****	21.807 (1.908)***
Sp. K. Working Memory	0.158 (0.008)***	0.018 (0.007)**
Fall K DCCS	0.190 (0.045)***	0.099 (0.029)**
Fall K Attentional Focus	0.637 (0.149)***	0.232 (0.079)**
Fall K Inhibitory Control	0.284 (0.13)*	-0.093 (0.067)
Fall K EBRS	0.080 (0.046)	-0.018 (0.017)
Fall K Full-Day	0.901 (0.438)*	-0.569 (0.173)***
Parental Pre-K Care	-0.195 (0.357)	0.119 (0.150)
Age at Kindergarten Entry	0.171 (0.039)***	-0.092 (0.016)***
Not First-Time Kindergartener	1.128 (0.648)	-2.509 (0.406)***
Socioeconomic Status	0.925 (0.196)***	0.468 (0.106)***
Non-Hispanic	0.111 (0.386)	-0.128 (0.172)
Female	-0.988 (0.219)***	-1.238 (0.138)***
Fall K Mathematics IRT	0.678 (0.019)***	-0.006 (0.017)
R^2	.81/**.70**	.21/**.19**
RMSEA	0.048	
SRMR	0.041	
BIC	172144.972	
χ^2 (*df*)	822.864 (120)	

Note. Bolded coefficients represent values for ELLs. *p < .05, **p < .01, ***p <.001

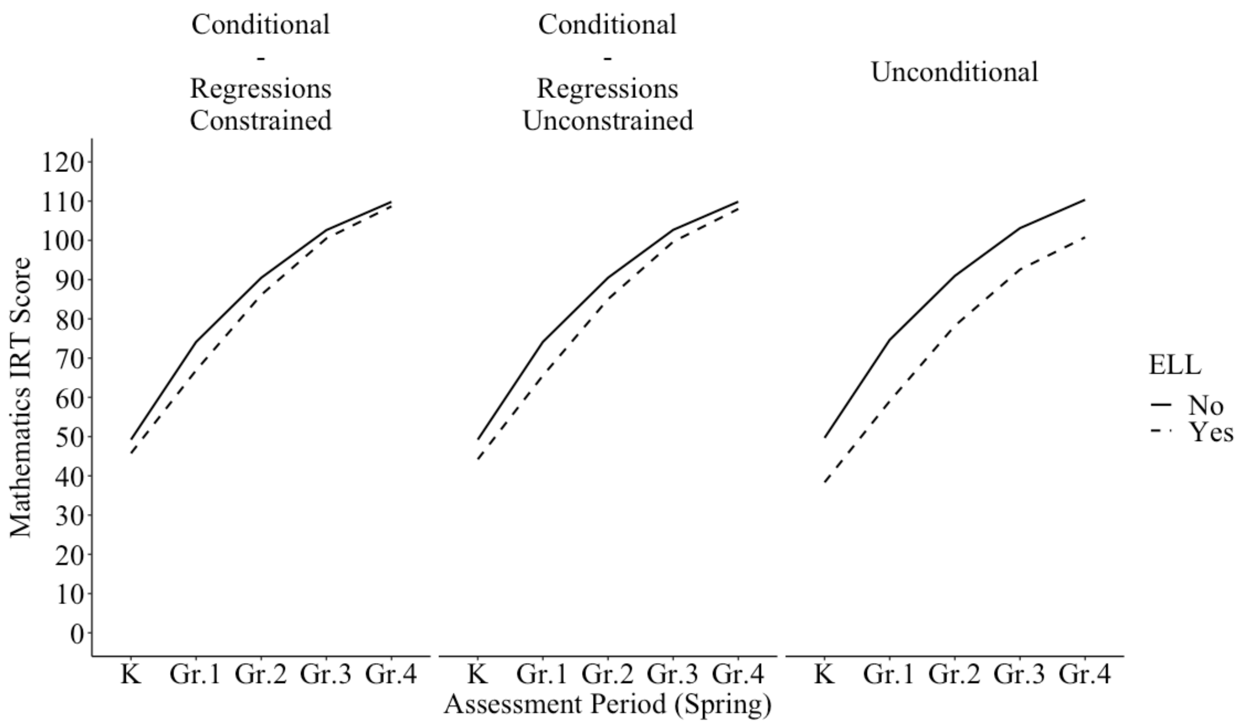

Figure 11. Estimated latent trajectories for each MG-LCSM by ELL status.

Note. The conditional trajectories are modeled at the means of each covariate within each group (including dummy codes) except Fall kindergarten mathematics scores, which are held at 0 across groups (i.e., the overall sample mean since the models were grand-mean centered). Figure produced in *ggplot2* (Wickham, 2016).

To answer Research Question 2, I allowed the regressions of the MG-LCSM intercepts and slopes on working memory to vary across groups while constraining all other regressions to equality across groups. This allowed me to test whether ELL status moderates the relationship of working memory in the MG-LCSM controlling for all other between-group differences in covariates. Freeing the working memory slope did not improve the overall model fit, as indicated by an increase in the BIC from 172144.972 to 172161.692. When freed across groups, working memory showed a significant relationship with both intercept ($b = 0.158$, $SE = 0.008$, $p < .001$) and slope ($b = 0.018$, $SE = 0.007$, $p < .01$) among EPSs. ELLs evidenced similar patterns, though only the working memory relationship with the intercept was significant (intercept: $b = 0.148$, $SE = 0.033$, $p < .001$; slope: $b = 0.015$, $SE = 0.012$, $p = .180$). This slightly worsened relative model fit suggests that working memory, holding other covariate effects constant across groups, likely does not differently predict mathematics performance level or change trend. In terms of predictive capability and parsimony, the MG-LCSM with working memory constrained to equality across groups is the preferred model. However, given that the BIC difference between the models is not substantial, this is not conclusive evidence that working memory operates exactly similarly across groups. The relationships with latent slopes are similar across groups, but the standard error among ELLs' is nearly double that of EPSs'. It is clear there is far more power to detect the effect when it is assumed to not vary across groups given the precision of the effect among EPSs due to the large sample size. This, coupled with the highly similar regression coefficient across groups, explains the BIC advantage in the constrained (i.e., unmoderated) model.

Study 1 Discussion

In Study 1, I used multiple group latent change score (MG-LCSM) to test the equality of change patterns across ELLs and their English-proficient peers and the extent to which language group moderated the role of working memory in mathematics development. The MG-LCSM framework permits direct tests of differences in additive change between kindergarten and fourth grade in addition to group differences in the patterns of change between individual Spring-to-Spring intervals. Detecting heterogeneity in the profiles of academic skills and growth over time is a core component of RTI as these differential change patterns help inform the extent to which different levels of performance relate to subsequent change across different developmental levels. Intraindividual change is integral to monitoring instructional response and forms the basis for curriculum-based measurement (CBM), which is why latent change score modeling can uniquely capture elements of academic growth that other modeling approaches do not (Petscher et al., 2016). Although it is not possible to extend the IRT scores of the mathematics measure in this study directly to CBM applications, these results provide a general picture of change among ELLs relative to EPSs and when fluctuations in change would be expected beyond the additive growth that occurs across early and intermediate elementary schooling. Nevertheless, given the nested nature of LCSM within the larger growth modeling approach, these results are similar to findings from previous work using the ECLS-K (Roberts & Bryant, 2011). Though the LCSM models similar patterns it simultaneously captures more information about the intraindividual change that explains what might be observed in a quadratic curve (i.e., where change occurs and the individual differences in performance levels that predict acceleration/deceleration).

In both groups, a slowing of growth was evident from kindergarten through fourth grade. This is consistent with prior growth modeling literature that has shown a plateauing of growth in mathematics through fifth grade (Grimm et al., 2010; Roberts & Bryant, 2011) among a variety of subgroups and through eighth grade among students with SLD in mathematics (Kohli et al., 2015). This study adds to this literature base by further quantifying and testing the significance of that plateauing via proportional change. However, the results of this study did not strongly support the hypothesis that ELLs would plateau more significantly than their English proficient peers. EPSs generally show more curvature to their growth pattern than ELLs, who from kindergarten to grade 3 show essentially a linear growth pattern. Relative to that linear change, ELLs show the hypothesized plateauing across third grade; however, this does not appear pronounced enough to suggest that ELLs change in a systematically different manner than EPSs in that area. Rather, EPSs appear to make greater gains from kindergarten to first grade, then there is deceleration in growth through the end of fourth grade (relative to the change that occurred in from kindergarten to first grade). The reason for this deceleration may in part be attributable to the distribution of scores becoming more left-skewed as grades progress.

Another aspect of these patterns is the substantive nature of the gains. The IRT-based scale scores used in the current study are comprised of the summed predicted probabilities of answering each item within the whole pool of items administered across all waves of data, and these predicted probabilities take into account factors such as students' latent ability level and the difficulty of the item (Najarian et al., 2018a, 2018b). Thus, growth along the IRT measure indicates a greater probability of answering more items correctly within the entire item pool administered to students in fourth grade. However, because the mathematics measure is comprised of multiple domains of mathematical knowledge with varying levels of difficulty, gains at different points on the scale are qualitatively distinct (Tourangeau et al., 2018). ELLs began and ended kindergarten performing nearly one standard deviation below their English-proficient peers. The gains that are then made in subsequent time intervals are significantly different from the gains of EPSs who begin kindergarten with higher levels of mathematics skills. Higher performing students may learn and expand their knowledge of mathematical concepts and procedures in different ways than ELLs given their potentially increased access to instruction. To that end, the skills that ELLs accumulate over the course of elementary school may be somewhat distinct from the skills initially higher performing students acquire. This represents a significantly different inference and implication than CBM measures would provide, which measure the fluency in a specific skill such as mathematics computation. Progress monitoring and growth in such a measure then represents growth in a specific skill domain of mathematics. The mathematics IRT measure, on the other hand, represents students' composite growth along a vertical scale of latent mathematics ability, taking into account item difficulty and multiple domains of mathematical concepts, facts, and procedures.

At the end of kindergarten, an average EPS may have gained significant competency in numbers and operations and thus may have a higher predicted probability of answering more difficult items on measurement or geometry. Subsequently, their gains across the following year through first grade are indicative of gains in more complex areas of mathematical knowledge (i.e., their summed predicted probabilities of answering more difficult items increases). On the other hand, ELLs exhibit lower predicted probability of correctly answering more complex mathematics problems. The gains that they make by the end of first grade, then, catch them up to the level at which EPSs were performing at the end of kindergarten. In other words, ELLs end kindergarten an entire grade level behind their English proficient peers. Differences in the estimated change patterns between the two language groups, as aforementioned, may be partly due to the shifting distribution of EPS mathematics scores across grade levels (e.g., greater left skew in later grades). Except for a small number of extremely low performing students, the distribution of ELL mathematics data is largely consistent across grades.

The lack of differential plateauing may be attributable to these qualitative aspects of academic gain. ELLs enter into a linguistically demanding mathematics environment performing at a level that allows them to continue making consistent, linear gains because the gains that are made are among more basic mathematical concepts, procedures, and facts. EPSs, on the other hand, continue to make gains in more challenging areas of mathematical knowledge. Though they decelerate in their growth, this indicates that it becomes more difficult to acquire the same knowledge gains as their level of performance increases and they are assessed in more difficult areas of mathematics. For example, EPSs exhibit a change score of nearly 25 points from kindergarten to first grade based on the MG-LCSMs, yet increasing by that same amount from third to fourth grade (relative to their prior level of performance) would be unlikely because that the 25-point gain in kindergarten is reflective of progress in basic number competencies and operations starting from a lower level of performance. The distribution of items in the kindergarten through second grade rounds of data collection weighs heavily toward number properties and operations; by third grade, these items comprise only forty percent of the items, with measurement, geometry, data and probability, and algebra gaining greater weight in the item pool (Najarian et al., 2018b). ELLs exhibit slightly greater change scores after kindergarten, which are likely reflective of greater room to grow because they do not reach as much of a ceiling in the difficulty of the items. Yet the gains they make in one year are the gains EPSs make in the prior year, suggesting ELLs are one year behind in the difficulty of items they are answering. Therefore, the similar decreases in gains by third grade observed across groups suggest that ELLs on average show decelerating gains in mathematical third grade mathematical knowledge during fourth grade instruction. Although the performance gaps do not increase during this time, it indicates that ELLs potentially decelerate in acquisition of content that their peers acquired a year earlier. Absent of higher dosages of effective, targeted intervention, this likely has the pernicious effect of diminishing their response to late-elementary and early middle school instruction in increasingly complex word problems, fractions, and algebra relative to EPSs.

Unsurprisingly, controlling for kindergarten-entry mathematics skills explains nearly all of the estimated growth gaps, suggesting that if ELLs and EPSs entered kindergarten with the same level of mathematics ability, their trajectories between the end of kindergarten and fourth grade would be highly similar (see Figure 11). Though some variance between groups is left to be explained, it is clear that kindergarten-entry domain-specific skills contribute the most to mitigating average achievement gaps over time. However, this does not change the underlying shape of development in either group.

Working Memory Findings

Controlling for other executive functioning, background, demographic, and academic (basic English reading skills and prior mathematics) variables, the finding that working memory continues to play a unique role in predicting mathematics development patterns for both ELLs and EPSs underscores its role in facilitating mathematics cognition across development. Prior studies, such as Cragg et al. (2017), highlighted its function in uniquely predicting procedural, conceptual, and factual mathematics knowledge, and Geary et al. (2017) showed that working memory is consistently a meaningful predictor of mathematics development across grades. To that end, in the current findings, early levels of working memory not only predict concurrent level of mathematics performance; early levels compound over time in the form of increase additive change over controlling for its effect on students' intercept (i.e., Spring of kindergarten). It is also important to highlight that working memory predicts these patterns net of students' kindergarten entry mathematics ability. Given the current results also show stark mathematics performance gaps at kindergarten entry, the current results on working memory suggest that working memory has a unique role in explaining mathematics development regardless of students' kindergarten entry mathematics performance levels. Another important aspect of this analysis to note is that this effect emerged controlling for kindergarten-entry English basic reading skills. ELLs scored significantly lower on average on the EBRS measure in the Fall of kindergarten (standardized mean difference [*SMD*] = 0.98, controlling for SES, race/ethnicity, sex, not first-time kindergartener, prekindergarten care, and age of kindergarten entry), and the contribution of working memory to growth patterns was similar across groups. Taken together, these results corroborate prior research establishing the critical role of verbal working memory in mathematics development (Cragg et al., 2017; Geary, 2011; Geary et al., 2012; Geary et al., 2017; Greenfader, 2017; Wang, 2017) while extending these findings to show that verbal working memory functions similarly across EPSs and low ELP ELLs net of other covariates previously shown to be strongly predictive of mathematics performance.

In lieu of a direct specification of empirical mechanisms that drive change among ELLs, the current findings are consistent with what has been identified as contributors to mathematics development among EPSs. First, working memory is a unique predictor of mathematics growth patterns both in terms of level and degree of gain, which is to say that students with higher working memory are simultaneously predicted to initially perform greater and subsequently make greater gains in mathematics through fourth grade. As I underscored throughout the literature review, working memory's developmental relationship to mathematics has a substantial body of empirical evidence. Despite the inability to assess differential relationships in specific domains of working memory (i.e., verbal versus visuospatial) in the current work, the results presented herein corroborate evidence of the meaningful role of working memory and lend further credit to such by controlling for a variety of other variables associated with achievement and subcomponents of executive functioning. By establishing that working memory does not differentially relate to mathematics across language groups, the current work extends prior findings to show that working memory is an equally important leverage point regardless of ELP. Both Fuchs, Schumacher, and Sterba et al. (2014) and Powell et al. (2017) showed that working memory ability evidenced strong capability to predict intervention responsiveness. Findings here of working memory's relationship to development suggest that it may play a similar, unique role among ELLs, though this requires scrutiny in an intervention context.

Recent work exhibited minimal incremental validity of cognitive variables in universal screening systems (Clarke et al., 2018), though these findings continue to align with evidence that cognitive variables explain unique variation in mathematics skills. The current study shows that response to instruction and consequent performance levels are not entirely domain-specific; regardless of kindergarten entry mathematics ability and basic English reading skills, working memory uniquely predicts mathematics performance levels and subsequent growth rates among both groups. Especially for ELLs, this underscores the crucial role that domain-general skills play in development net of domain-specific performance in mathematics and reading. Given that ELLs begin and end kindergarten performing an entire grade level behind EPS peers and continue to evidence this gap through fourth grade, working memory's unique contribution to developmental levels and trends points to the need for building domain-specific skills while leveraging domain-general areas to support learning regardless of second language development.

However, as Clarke et al. (2018) also note, introducing domain-general variables into screening needs to be balanced with time and economic value of doing so. The current findings warrant further investigation into which domain-general skills are most uniquely predictive in order to better clarify the incremental value of such skills in applied settings.

Assuming that second language learning may impact the relationship of an executive functioning component like working memory and mathematics may be an easy conclusion to draw; however, these exploratory findings indicate otherwise. This is not to say that language does not play a role in mathematics; as aforementioned, it is a critical component of multiple aspects of mathematical cognition and instructional response. However, further investigations ought to focus on the mediators between working memory and mathematics, as the process of this relationship may heavily rely on the extent to which working memory helps build mathematical language in L1 and L2 and how that transfers to building mathematical knowledge. Indeed, using the ECLS-K: 2011, Wang (2017) found that higher growth rate in working memory through grade three mediated the effect of kindergarten ELP status on mathematics growth through grade three. Non-proficient bilingual students in Wang (2017) scored significantly lower on working memory, possibly leaving greater room to grow over time on that measure. Similarly, ELLs in this current study scored significantly lower on working memory in the Spring of kindergarten (b = -6.83, SE = 1.20, t = 5.71, SMD = 0.42, p < .001) even after controlling for first-time kindergartener status, age of kindergarten entry, race/ethnicity, sex, taking the cognitive assessments in Spanish, SES, and pre-kindergarten care. With this gap in mind, such increases in working memory likely facilitate the growth of mathematical content knowledge, as increasing executive functioning skills like working memory may help students access curricular content and respond to instruction more effectively. This may especially important when ELLs end kindergarten scoring nearly half a standard deviation below age-based norms (even controlling for demographic factors). A significant gap such as this warrants further empirical investigation of mechanisms, as the processes through which ELLs may access instructional content and generate mathematical knowledge may operate through different mechanisms if they begin with significantly lower working memory.

That said, although these results also point towards the notion that although ELLs observationally exhibit many different risk and protective factors in their learning, these do not explain away the more basic aspects of how executive functioning contributes to mathematics learning. When students' native language (and the language of instruction) can be effectively accounted for, the predictors of students' learning appear similar. To that end, identifying the most equitable and effective ways to instruct mathematics while providing as few language barriers as possible should continue to be a critical focus of research in educational policy, psychology, and curricula.

Limitations

The current study has significant limitations that are crucial to address in light of the findings. Principally, my definition of ELLs may not be consistent with the definition schools use according to summative measures of ELP. These assessments provide a complex and thorough assessment of English language proficiency – both academic and social – that cannot be equated with the definition of ELL I used in the current work. Although the *pre*LAS provides a brief screening assessment of language proficiency and parent-reported native language has been utilized in prior work (e.g., Roberts & Bryant, 2011), this definition may not easily extrapolate to the students classified as ELLs in schools based on state-mandated, IRT growth measures of ELP such as ACCESS 2.0 (World Class Instructional Design and Assessment, 2014).

Additionally, the current study utilized mathematics and executive functioning assessments in which a number of students were assessed in Spanish. In kindergarten (Fall and Spring), this constituted a significant proportion of students. While the IRT scale of mathematics is consistent across Spanish and English, this may represent a confound in comparing growth across ELLs and EPSs and the ability to translate this information into practice in an RTI model. Mainly, many schools assess ELLs in their native language (as of now, mostly Spanish assessments exist) and in English. In this way, schools may address academic progress simultaneously in both languages and use this information to determine the need for further assessment, tiered intervention, and services in ESL programs. With some students having been assessed in Spanish, I am limited in my ability to parse-out the role of language in students' mathematics growth through the end of first grade. Relatedly, some ELLs' native language was likely one other than Spanish, though I do not have access to this information in the public-use ECLS-K: 2011. Non-Spanish speaking ELLs who scored below the *pre*LAS cut score did not take any of the cognitive assessments; thus, these students were removed from the data. Prior studies have utilized parent- and school-reported race/ethnicity as a proxy for native language (e.g., Roberts & Bryant, 2011); however, it was not the goal of the current study to analyze differences based on different native language. Students whose native language was not Spanish nor English would have been assessed in English (if they scored 16 or above on the *pre*LAS), making it more difficult to precisely measure mathematics growth within the ELL group given the lower-end of the ELP scale among non-Spanish-speaking is not at all represented. Hispanic/Latino students also represent the overwhelming majority of ELL students in the current sample (> 90% of the sample), hence the non-Spanish-speaking ELLs are not adequately represented. Though Spanish-speaking students represent a majority of ELLs in the U.S., the current demographic composition of our sample likely does not represent the general distribution of students in U.S. public schools. Yet more research is needed to understand patterns of achievement among non-Spanish-speaking ELLs, as the cognitive and cultural factors implicated in their instructional response may vary significantly.

Another limitation is the comparability of ELL and EPS groups. First, the distribution of SES between the two groups is not comparable. This is a frequent limitation of studies involving comparing ELLs and EPSs (Abedi & Lord, 2001), as SES is typically associated related to achievement patterns. It is generally understood that language accounts for variance in achievement at varying levels; however, the relative contribution of SES to these achievement patterns in combination with language is less clear. For example, Abedi and Lord (2001) determined the SES and second language proficiency among ELLs were confounded. The goal of the current study was to translate differential achievement patterns into findings to inform RTI practices for ELLs, not to directly model the role of SES in ELL mathematics achievement patterns, though it is understood that SES is a general predictor of academic development, and the role of SES in development can play unique roles of ELLs given the diverse demographic backgrounds of students and their families (e.g., Reardon & Galindo, 2009). Nevertheless, this difficulty in distinguishing between the role of SES, other contextual factors (e.g., time spent in U.S.)., and language proficiency continues to limit the conclusions of this research. The confounding of SES and language proficiency (Abedi & Lord, 2001) makes it highly difficult to determine if the between-group differences that disappear after controlling for SES across groups is then due to the fact that SES explains most of the between-group differences or because controlling for SES inherently controls for language differences. This recurring difficulty is something that needs to be addressed by studying more within-group heterogeneity among ELLs to determine when and among whom SES is most predictive of achievement patterns. The current study could not adequately parse out these relationships (e.g., ELLs were nearly 1 *SD* below EPSs on kindergarten SES), nor was it the goal to explicitly study SES. Relatedly, from a practitioner stand point, considering SES itself as a risk factor is not in the direct purview of the RTI model. However, ecological approaches to working with students from a range of socioeconomic strata, such as accessible and equitable family engagement practices to support academic skills and also social-emotional-behavioral well-being, may be effective in mitigating the roles of exogenous variables in students' academic, cognitive, and social-emotional-behavioral development.

Finally, one frequent confound to modeling growth is the variability in when the outcome is observed across waves. In the ECLS-K: 2011, Spring assessments took place in March through June of each year. As such, the intervals between measures for some students could have been as few as nine months or as much as 15 months. One significant disadvantage of the MG-LCSM approach in this area is that a basic assumption of the LCSM is equal intervals of measurement (McArdle & Nesselroade, 2014). Potential ways to mitigate these problems might include using a placeholder latent true score that is estimated from the model between each latent true score estimated from the actual data (McArdle & Nesselroade, 2014), which would then generate an intermediary interval estimated from the model and may capture variation in timing of observations. Although a core assumption of the LCSM approach, this issue has not been evaluated in the literature, making it difficult to assess the impact individually varying times of observation has on model estimation. Nevertheless, the model results and the assumption of a yearly change period should be assessed in light of this fundamental limitation of the data.

CHAPTER IV: STUDY 2

Research Questions and Hypotheses

The mains goals of this study were to, first, determine the extent to which gains in ELP across kindergarten predict mathematics development patterns between grades 1 and 4 when the majority of students have been assessed in math in English and, second, model potential unobserved heterogeneity in mathematics development patterns between first and fourth grades. Three primarily exploratory research questions are the focus of Study 2.

Research Question 1

Are there unobserved differential growth trajectories (i.e., mixtures) in mathematics development from grades 1 to 4? Even though average ELL performance is lower on average than EPS performance, do these potential heterogenous trajectories align with the average status/steeper growth ("mover") lower status/flatter growth ("stayer") trajectories commonly identified in other mixture modeling findings across language groups (e.g., Kaplan, 2002; Jordan et al., 2006; Jordan et al., 2007)? If so, to what extent do these growth mixtures align with current assumptions about RTI tiers of defining academic risk (Fuchs & Fuchs, 2007)?

Research Question 2

Do gains in ELP across kindergarten predict greater mathematics growth? If multiple growth classes exist and align with the mover/stayer patterns of development, does gain in ELP across kindergarten predict membership to those classes and the probability of switching classes? Do ELP gains function differentially across mixture classes in predicting of growth patterns? Prior findings show the timing of reaching English proficiency is predictive of mathematics development (Halle et al., 2012), so I expected ELP gains to be positively related to later mathematics development trends given that greater gains imply greater progress towards ELP over kindergarten and thus improved access to instruction and assessment. However, prior research has focused largely on level of ELP rather than gains in ELP as a predictor of academic development, though ELP growth is predominantly the focus of policy and practitioner initiatives. Due to the paucity of research relating ELP growth and mathematics, these analyses were largely exploratory, though I did primarily expect that ELP growth would positively predict later mathematics development trends. However, it could also be the case that greater gains in ELP early on are not necessarily beneficial, especially as the students who gain more in ELP are likely to be lower in ELP to begin with. Chow et al. (2013) found that greater proficiency and gains in early reading may drag behind mathematics development and noted this could be related to the disproportionate amount of time students potentially spend on acquiring literacy skills in early elementary school. Schools may feel that it is more imperative for students to progress in ELP and literacy as the more fundamental academic skill, though this may come at the expense of mathematics development to some degree. Nevertheless, given the prior findings indicating the importance of the timing of ELP, the prediction leaned more heavily in favor of ELP gains positively predicting later growth. However, if growth mixtures can be extracted, it is possible the role of ELP may function differently depending on the qualitative characteristics of such mixture distributions. For example, Chen and Chaloub-Deville (2016) employed quantile regression to show that English reading ability was more strongly predictive of mathematics performance among students with lower mathematics performance. Similarly, it might be the case that ELP gains are more strongly predictive of later mathematics development among students who may exhibit riskier mathematics trajectories.

Research Question 3

Because I primarily predicted that ELP growth would be a positive predictor of later growth, it was also important to identify differential relationships in the extent to which ELP predicts later growth. Moderators can be particularly useful in the context of RTI given the emphasis on student characteristics that predict responsiveness to intervention and the need to assess profiles of mathematics and related skills to identify such responder patterns (Fuchs, Schumacher, Sterba, et al., 2014; Powell et al., 2017). ELP might play a role in intervention response (e.g., Doabler et al., 2019), though recent evidence (Swanson et al., 2018) clearly establishes that ELP does not tell the whole story in terms of the mechanisms that drive mathematics development among ELLs, and by extension, potentially intervention response. Therefore, Question 3 probed whether ELP is a stronger predictor of mathematics trajectories among students with lower working memory capacity. If multiple classes of growth exist, how does this interaction function across growth classes? ELP coupled with needing to manipulating multiple linguistic sources of information in mathematics (i.e., concepts, facts, and information partially stored and retrieved in native language but applied in a second language task), lower working memory capacity may increase the load of language in mathematics tasks because lower working memory may make mathematical problem solving less efficient. Prior work has established mixed findings on the role of different components of executive functioning in terms of bilingualism; however, working memory would seem to most effectively link L1, L2, and mathematics because of the role of working memory in mathematics over time and the potential need of some (but not all) ELLs to attend to, hold, and manipulate a wider variety of information sources than EPSs. Indeed, Swanson et al. (2018) provided evidence for this critical role of working memory among high and low ELP ELLs with and without mathematics difficulties. As such, ELP may be less predictive of mathematics development at higher levels of working memory because those students might more effectively utilize their working memory resources to offset some of the language load of mathematics problems, thus gaining more in ELP may not be as important in predicting later performance. Students with lower working memory ability may be less effective in utilizing executive functioning resources to work through mathematics problems with more language and cognitive load, thus the mathematics task could become more contingent on their level of ELP. Consequently, early ELP gains may be more critical to later mathematics performance and development for these students. If mixtures are present, then this interaction might function differently for students within each mixture.

Study 2 Method

Participants

Participants for Study 2 were the ELL subsample utilized in Study 1 (unweighted N = 303; weighted N = 205,956).

Measures

The same measures used in Study 1 were used in study two except the ELP measures, which I describe below.

English language proficiency growth. I assessed growth in English language using Fall to Spring gains in kindergarten on the *pre*LAS, which models the latent difference between the Spring and Fall measurements of the *pre*LAS. I defined ELLs in this study in the same way as Study 1. Thus, I measured gains from Fall to Spring among students who scored between 0 and 22 on the *pre*LAS in the Fall. Fall and Spring α coefficients of the *pre*LAS total raw scores (i.e., total score of 20) were 0.91 and 0.91, respectively. As a secondary related but distal measure of English proficiency growth, I also utilized the English Basic Reading Skills (EBRS) portion of the ECLS-K: 2011 reading assessment. This measure consists of 18 basic reading skills items in addition to two items from the *pre*LAS "Art Show" task (Tourangeau et al., 2018). The content covered on the EBRS in kindergarten included phonological awareness, print familiarity, recognition of letters and sounds, and sight words (Najarian et al., 2018b), whereas the *pre*LAS measured only expressive and receptive vocabulary in English. There is likely considerable overlap in the underlying language skills measured between the two besides the two common *pre*LAS items since proficiency in expressive and receptive vocabulary likely shares similar characteristics as L2 skills such as phonemic awareness and letter-sound recognition.

Missing Data

With the exception of the treatment ELP data, all other variables used in Study 2 were the same as Study 1. As such, the same set of 20 imputed datasets generated in Study 1 was utilized in the current study.

Analysis

I used the SUBPOPULATION command in M*plus* to analyze the subset of 303 ELL students from the larger dataset used in Study 1. Because the data are weighted to produced population estimates (using the same sample weight, PSU, and strata as Study 1), traditional subgroup analysis that fully removes data from an analysis may not produce the correct standard error estimates. To account for this, the SUBPOPULATION command utilizes data from the entire dataset but weights the group that is not being used in the analysis (in this case, EPSs) to 0. This allows the full sample to contribute to standard error calculation while providing estimates for only the weighted subgroup of interest.

The analytic technique of this study consisted of multiple exploratory stages. First, I employed GMM[3] to detect heterogeneity in growth trends between the Spring of grade one and the Spring of grade four, totaling four measurement waves. Provided that a multiple-class mixture model fits better and is substantively more meaningful than a single, latent curve model, the second step of the modeling process is to include covariates into the model using the three-step method (Asparahouv & Múthen, 2014). Because mixture modeling takes into account all information that is included into a model during the estimation process, including covariates into the entire model may substantially impact the estimation of mixture classes in the outcome variables. Although in some cases this may prove empirically and substantively advantageous (Stegmann & Grimm, 2018), the primary goal of mixture modeling is to detect heterogeneity in the outcome variable rather than in both the outcome itself as well as the covariate regressions (Jaccobucci, Grimm, & McArdle, 2017). The three-step method offers a solution to this challenge by classifying the covariates as auxiliary variables in the estimation model, which effectively removes the covariates from the extraction of mixture classes but then allows class membership to be regressed on the covariates after the mixture class estimation procedure. The third step in this analysis procedure is testing the stability in mixture class membership and the factors that predict stability or change (i.e., regime-switching; Chow et al., 2013). The current M*plus* software (Version 8.1; Muthén & Muthén, 2017) does not have an automatic procedure for conducting the three-step method for regime-switching growth mixture models, as the three-step procedure can handle identification of only one set of latent mixture classes. Consequently, I will use a latent transition analysis (LTA) technique to analyze stability and change in class

[3] Unlike Study 1, individually-varying times of observation can be more easily included into the estimation of a growth model. However, this requires using the TYPE=RANDOM command, which would then preclude the use of both mixture modeling and complex design features. As such, individually-varying times of observations were not permissible in Study 2.

analysis membership. To conduct this technique, I will conduct a latent profile analysis (LPA) of each of the four mathematics measurements utilized in the GMM. Membership to the classes at each wave will then be extracted and combined into one data set. Then, an LTA will be conducted by regressing classes membership to membership at the previous time point. Predictors can then be included into the LTA model through the three-step specification technique identified in Nylund-Gibson, Grimm, Quirk, and Furlong (2014). Although this technique does not identify growth mixtures per se, the specification of classes at each measurement wave in the RS-GMM parallels the process of conducting LPA at each wave, extracting class membership for each wave, then regressing that membership on each prior wave. The LTA approach is advantageous because it estimates fewer parameters within the same model, offering a more parsimonious approach to the modeling process. Given the complexity of estimating the mover-stay component within a RS-GMM – which entails the specification of a GMM, the specification of the mixture classes at each wave, the regression of wave-specific mixture membership on the prior wave's membership, then the added model constraints to identify the "movers" versus "stayers" – the added simplicity of conducting the mover-stayer modeling within just the LTA is significantly advantageous and estimated with relative ease in Mplus (Kaplan, 2008; Kaplan & Walpole, 2005).

A primary aspect of mixture modeling is the selection and substantive identification of mixture classes based on qualitative and quantitative indicators. On the quantitative end, the BIC is frequent measures of mixture identification as it assesses model parsimony by assigning a penalty for estimating unnecessary parameters (Kaplan, 2002; 2009; 2014). Lower BIC values suggest better model fit (Kaplan, 2002; 2009; 2014). Mixtures have to make sense substantively, however. Over-extraction of mixture classes has been raised as a significant issue in the GMM literature (Bauer & Curran, 2003). Simply because separate mixtures of growth trajectories can be empirically identified does not necessarily imply that the mixtures are qualitatively and substantively meaningful. Thus, mixture class identification ought to be guided by both the data at-hand (i.e., relative fit indices like the BIC [Kaplan, 2002]) and an understanding of what underlies observed heterogeneity in growth trajectories.

The first step of growth mixture modeling is to descriptively inspect profile plots of individual growth trajectories (Kaplan, 2002; Ram & Grimm, 2009). Profile plots of a random sample of 30 students out of the 303 individuals in the current study are provided in Figure 12. Additionally, because of the interest in ELP gains in predicting mathematics growth, profile plots of data from the two ELP measures under consideration from 30 randomly sampled students are presented in Figure 13.

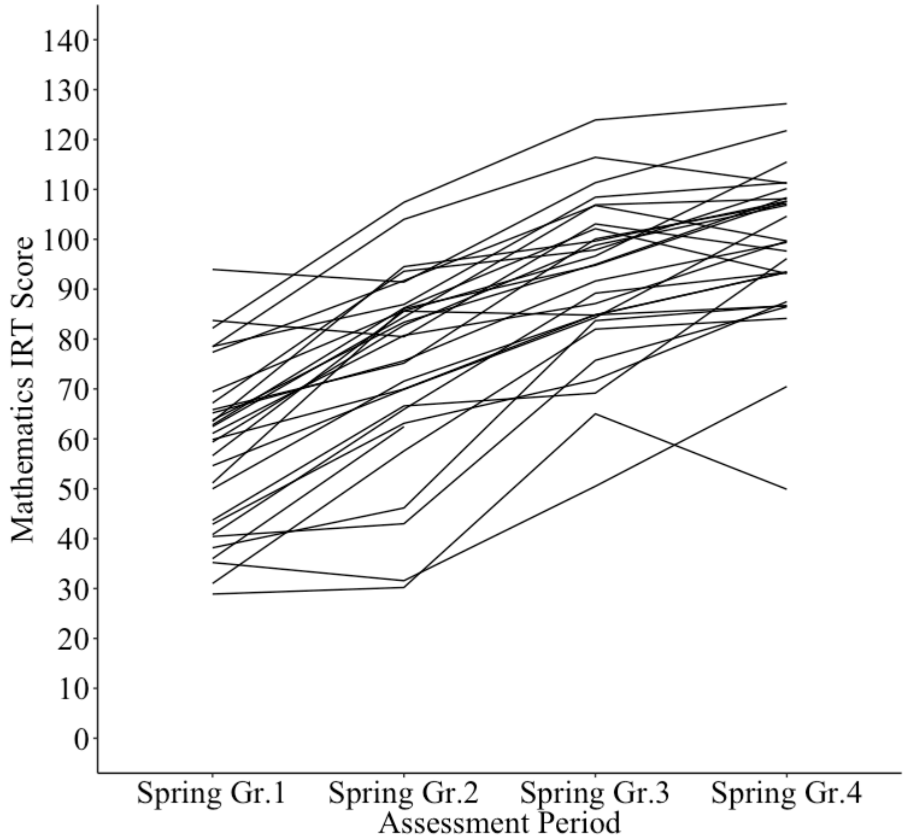

Figure 12. Profile plots of mathematics IRT scores in first through fourth grade for 30 randomly sampled ELL students.

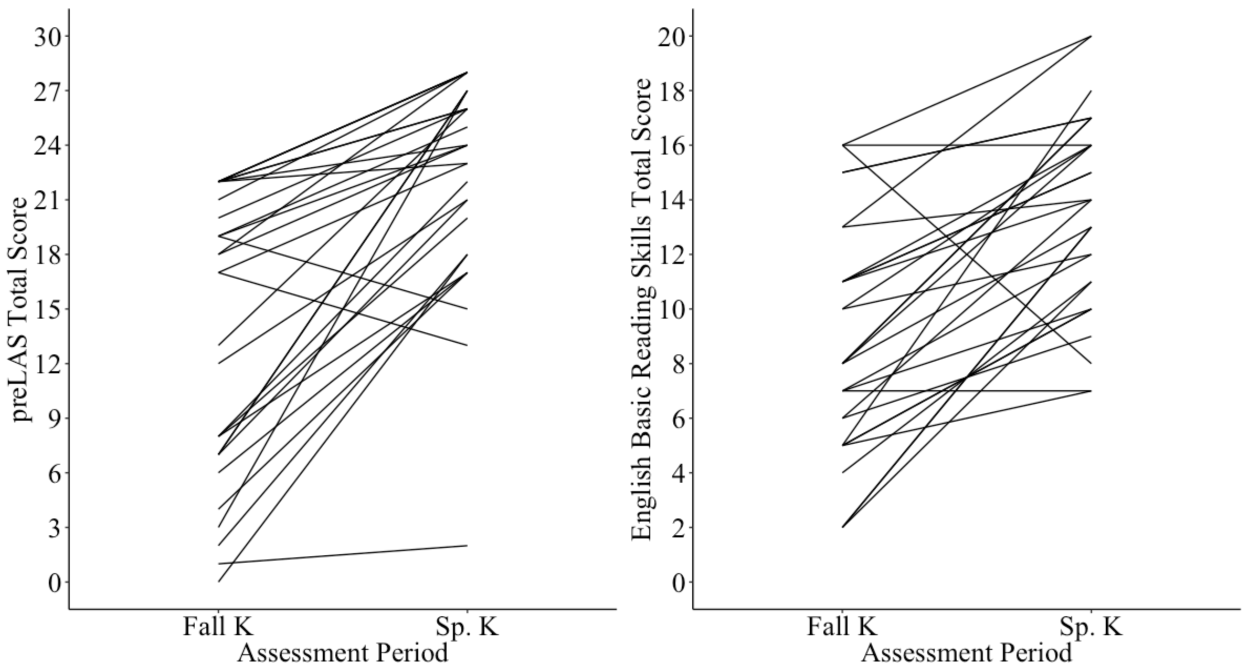

Figure 13. Profile plots of *pre*LAS and EBRS data in Fall and Spring of kindergarten from 30 randomly sampled ELLs

For the most part, students appear to exhibit relatively stable trajectories over time based on Figure 12. However, students who began the Spring of first grade with scores on the lower end of the distribution exhibit considerably more variability in their trajectories until the end of grade four, particularly in the one year between first and second grade. This variability in initial status in addition to the ensuing trajectories indicates that a single latent growth curve may not adequately capture the underlying growth trend as hypothesized; thus, GMM is a justifiable approach.

I estimated the GMM using an iterative procedure (Ram & Grimm, 2009). One-, two-, and three-class unconditional models were investigated first with means varying across classes, then with means and variances/covariances varying across classes, and finally with means and variances/covariances freed across classes with residual variances freed across time. I freed residual variances of manifest variables (i.e., mathematics measurements at each wave) over time within each class, as this may provide important information on the unaccounted variance (i.e., disturbance) within each mixture class. Variance that is not explained by students' initial status (intercept) or their pattern of growth and the extent to which this error changes over time may be potentially important information in qualitatively distinguishing growth patterns. Residual terms are frequently not of substantive interest nor clearly discussed in studies (Enders & Tofighi, 2008). However, Enders and Tofighi (2008) conducted a simulation study showing that analysts must attend to manifest residual terms in GMMs, as incorrectly constraining or estimating these terms can lead to over-extraction, bias, and misspecification. Thus, because the observed trajectories depicted in Figure 12 appeared heterogenous, I more closely scrutinized the viability of assuming the equality of residuals of time within each class. The reliability of slope estimates is an important aspect of assessing the technical adequacy of progress monitoring tools (Christ &

Ardoin, 2008; Clarke et al., 2008), and higher unexplained variance may be indicative of less reliable estimates of growth. The SEM growth modeling framework relaxes the homoscedastic residual assumption and offers an approach to estimate the variance for which students' latent trajectories do not account.

Study 2 Results

Descriptive statistics for all the variables used in this analysis are presented in Table 13, and demographic characteristics of the ELL sample are presented again for reference in Table 14.

Table 13

Study 2 ELL Weighted Sample Descriptive Statistics

Variable	*M*	*SD*	Min	Max	Proportion Missing
EBRS – Fall K	8.05	4.63	0.00	20.00	.000
EBRS – Sp. K	13.65	4.21	0.00	20.00	.000
IRT Mathematics - Fall K	23.73	7.60	9.79	48.54	.000
IRT Mathematics – Sp. K	38.24	10.58	14.13	76.11	.000
IRT Mathematics - Sp. Gr. 1	58.93	13.88	25.88	103.40	.000
IRT Mathematics - Sp. Gr. 2	78.10	16.02	23.45	115.43	.000
IRT Mathematics - Sp. Gr. 3	92.66	15.14	43.95	123.92	.001
IRT Mathematics - Sp. Gr. 4	100.87	14.77	28.16	130.95	.001
WJ - III Numbers Reversed – Spring K	85.22	16.46	51.00	134.00	.000
DCCS – Fall K	12.09	4.12	0.00	18.00	.000
CBQ Attentional Focus – Fall K	4.39	1.32	1.00	7.00	.063
CBQ Inhibitory Control – Fall K	4.73	1.32	1.00	7.00	.056
Age at Kindergarten Entry	65.02	4.83	39.10	81.07	.003
Socioeconomic Status – Fall/Sp. K	-0.84	0.54	-2.33	1.61	.003
*pre*LAS Combined Score - Fall K	13.56	7.57	0.00	22.00	.000
*pre*LAS Combined Score – Sp. K	22.04	6.74	0.00	30.00	.000
Unweighted *N*			303		
Weighted *N*			205,956		

Note. Missingness proportions are not weighted. EBRS = English basic reading skills, IRT = item response theory, WJ = Woodcock-Johnson, DCCS = Dimensional Change Card Sort, CBQ = Children's Behavior Questionnaire, K = Kindergarten.

Table 14

Study 2 Weighted Demographic Characteristics (Proportions)

Variable	Proportion	Proportion Missing
Black	.007	.000
White	.006	.000
Hispanic, Race Specified	.91	.000
Hispanic, No Race Specified	.01	.000
Asian	.05	.000
Hawaiian/Native Pacific Islander	.01	.000
American Indian/Alaskan Native	.002	.000
Female	.50	.000
Not First-Time Kindergartener	.05	.003
Parental Pre-K Care	.38	.003
Attended Full-Day Kindergarten (Fall)	.91	.007
Attended Full-Day Kindergarten (Spring)	.93	.000
Unweighted *N*	303	
Weighted *N*	205,956	

Note. Missingness proportions are not weighted.

To address Research Question 1, I conducted GMM in M*plus* (Muthen & Muthen, 2017) with the TYPE = COMPLEX MIXTURE command, which allows the incorporation of complex survey design features (weights, primary sampling units, and strata) into the mixture modeling procedure. I tested models in an iterative fashion starting with a single-class latent-basis growth curve model. I chose the latent-basis model (as opposed to a quadratic or linear) given the apparent overall nonlinearity in change patterns and the inconsistency of nonlinearity. Moreover, given the descriptive plots of individual curves, a latent-basis model seemed more capable of capturing differential patterns of change over time than a model with a strict functional form. Relatedly, due to this variability in change, I left residual variances unconstrained in the single-class model. Then, I fit a two-class mixture model with residuals constrained over time and variances/covariances constrained to equality across groups (i.e., only the intercept and slope means varied between groups). Following that model, I freed the latent intercept and slope variances/covariances. Last, in a third two-class model I freed the residual variances as well as the latent intercept slope variances and covariances. I conducted the same procedure with a three-class mixture mode. However, the two-class models with freed means and variances/covariances and residuals constrained across time (but freed across groups) showed a non-positive-definite matrix and was therefore removed from consideration. Similarly, the three-class model with the same constraints had a non-positive-definite matrix, as did a three-class model with completely freed means, variances/covariances, and residuals. As such, only four models were left for consideration in the class enumeration procedure.

Results of this analysis are presented in Table 15. BIC values from each mixture model (one-, two-, and three-class models) suggest a slight improvement in model fit from a one to two class model with freed means and a two-class model with all parameters freed across groups.

However, despite the lower BIC and entropy values were quite high (>.90), the classes were not qualitatively meaningful since the class mixing proportions were very small for some classes (≈.06 – .09). Although this range of proportions may mirror the population estimate of students who exhibit specific learning disorder in mathematics (American Psychiatric Association, 2013), there is not a specific way to determine this with the available data. Even though a binary variable indicates whether students have a disability or not, this would not provide any information regarding how their diagnostic classification might relate to mathematics outcomes. Additionally, in both the two-class models, the only significant quantitative differences observed between mixture classes was the intercept, the manifest residual terms (when they were freed), and one growth factor loading in the latent-basis model (grade 2). Otherwise, the trajectories closely modeled each other. The three-class model presented no advantage in terms of class distinctiveness (entropy) or relative model fit (BIC).

Data inspection among the ELL subsample from Study 1 detected a significant univariate outlier in the fourth-grade mathematics data. Although this value did not meaningfully impact prior analyses, it was further inspected here given that this particular student was included in the highly small mixture class. Mixture models were re-fit with this student removed from the models. Model fitting for single-class as well as the two- and three-class models with only freed latent means was unaffected by the removal of the outlier. The difference in BIC values between the single and two-class freed means model slightly shrunk, however, providing further evidence that the single-class model was likely the proper selection. The two-class model with freed means, variances/covariances, and residuals showed a non-positive-definite matrix, so it was removed from consideration. However, it is interesting to note that the removal of this outlier significantly impacted the already highly small mixing proportion, suggesting it was driving much of the mixture detection.

Table 15
Relative Fit Indices of Growth Mixture Models

	BIC	Entropy	Class Mixing Proportions*
Including Gr. 4 Outlier			
Single Class (freed residuals)	8980.38	--	1.00
2-Class: Means	8966.84	.83	.91, .09
2-Class: Means + Variances (Covariances) + Residual Variances	8973.23	.95	.94, .06
3-Class: Means	8980.51	.65	.09, .55, .37
Gr. 4 Outlier Removed			
Single Class (freed residuals)	8930.81	--	1.00
2-Class: Means	8922.23	.85	.92, .08
2-Class: Means + Variances (Covariances) + Residual Variances^	8934.31	.99	.97, .03
3-Class: Means	8933.85	.67	.09, .57, .34

Note. All residual variances estimated separately within groups. ^Model showed non positive-definite matrices.
*Based on estimated posterior probabilities.

Considering the minor empirical advantage to the two different two-class models, the substantive relevance of the classes, and the influence of the outlier in the mixture detection procedure, a single class (i.e., no mixture) growth model appeared to be the model that most accurately represented the growth trajectories. Although the weighted population mixing proportions correspond to class sizes of 12,063 for two-class model with means, variances/covariances, and residuals freed (outlier included) and 19,051 for the outlier-removed two-class model with only freed means, there is nonetheless unconvincing empirical support from a model parsimony perspective (i.e., the BIC) that specifying these mixtures provides a better model. Comparing these mixing proportions to the entire population (weighting the entire sample of 5,014 students), these mixing proportions constitute only between 0.4% and 0.6% of the overall population (e.g., 12,063/3,432,072 = .004). While the goal of the current analysis was to identify unique mixtures specifically among ELLs, it is important to examine these results in the broader context of the general population being studied. The potential to observe a fundamentally unique trajectory of mathematics growth within a unique subpopulation (i.e., ELLs) in the context of the entire population with class proportions of less than half a percent seems unlikely from a practical perspective. Of course, any proportion of mixtures among ELLs would inherently be quite small when compared to the whole population. However, the lack of qualitative distinctiveness and cogent empirical support among ELLs specifically provides greater evidence that it is unlikely these mixtures are substantively or practically relevant at a population level.

Generally, the single-class model can be characterized by high intercept and slope variance with high observed variable variance at each measurement wave. Similar to Bauer and Curran's (2003) cautions against over-extraction of mixture classes, this high variability in growth patterns does not by default mean mixture classes exist. Because only a single class model was selected, a regime-switching model or latent transition analysis was not permitted. Results of the single-class model are presented in Table 16. Removing the outlier had no significant impact on model estimation besides lowering the BIC, so it was retained in the model estimation.

Table 16
Unconditional Latent-Basis Growth Parameters

Parameter	Estimate (SE)
Latent Intercept Mean	58.91 (1.10)***
Latent Intercept Variance	175.33
Latent Slope Mean	13.98 (0.22)***
Latent Slope Variance	8.55
Intercept-Slope Covariance	-8.19 (4.68)^
Latent-Basis Slope Loadings	
Sp. Gr. 1	Fixed at 0
Sp. Gr. 2	1.38 (0.05)***
Sp. Gr. 3	2.41 (0.03)***
Sp. Gr. 4	Fixed at 3
Residual Variances	
Sp. Gr. 1	24.46
Sp. Gr. 2	53.70
Sp. Gr. 3	37.68
Sp. Gr. 4	22.14
Fit Indices	
RMSEA	0.112
CFI	0.991
TLI	0.967
SRMR	0.114
BIC	8980.38
χ^2 (*df*)	14.344 (3)

Note. To properly calculate CFI and TLI values for growth models (Widaman & Thompson, 2003), the CFI and TLI are calculated using an intercept-only model with residuals constrained to equality over time as the null model (χ^2 [*df*] = 1284.060 [11]). ^$p \geq .05$, ***$p < .001$

After identifying the best-fitting model (i.e., the single-class latent-basis growth model), I then included covariates to analyze the role of ELP development across kindergarten in predicting future trends in mathematics development. However, prior to that, latent change score models for Fall to Spring *pre*LAS and EBRS were estimated separately. Models were fit in R using the *lavaan* (Rosseel, 2012) and *lavaan.survey* (Obserski, 2014) packages to estimate the model with complex survey design elements. Robust maximum likelihood ("MLM", as opposed to "MLR" in M*plus*) was used in estimation. Path diagrams for these models are presented in Figure 14. Because no data were missing for the *pre*LAS and EBRS measures in either Fall or Spring of kindergarten, the models were fit using the original, single dataset. When regressing change on prior level (autoproportional effect), both models show similar autoproportional effects: students who performed higher on *pre*LAS or EBRS by one point gained 0.57 and 0.66 fewer points by Spring, respectively. Because the latent change score is constructed from observed variables (rather than multiple indicator variables), manually-constructed and latent change scores are mathematically equivalent (McArdle, 2009). Manually-constructed change scores were thus used in subsequent analyses. On EBRS, ELLs gained an average of 5.65 points, and on *pre*LAS students gained an average of 8.52 points. Another advantage of using manully-constructed change scores is that it avoids estimation barriers with latent interactions, which cannot be used with complex survey features in M*plus*.

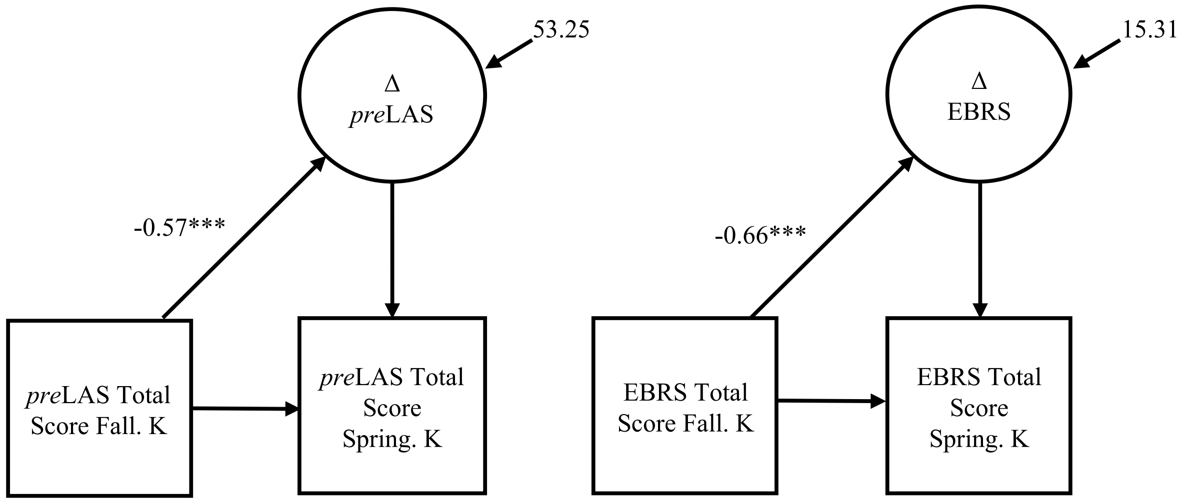

Figure 14. *pre*LAS and EBRS change score diagrams.

Note. Unlabeled paths fixed to 1. *** $p < .001$.

Next, the full conditional growth curve model was estimated with the full set of covariates. Table 17 displays the regression coefficients for each of the predictors in the model with *pre*LAS and EBRS as predictors, and Table 18 displays the latent growth parameters and fit indices for each of the two the conditional models. All continuous predictors were centered at the ELL-specific mean of each variable to aid in interpretation of the EBRS/*pre*LAS-by-working memory interaction term and the conditional EBRS and *pre*LAS effects.

Table 17
Covariate Effects in Latent-Basis Growth Curve Model

	Regression Coefficients (SE)	
Predictor	Growth Intercept	Growth Slope
***pre*LAS Model**		
(Intercept)	56.122 (1.340)***	14.967 (0.624)***
Sp. K. Working Memory	0.239 (0.058)***	-0.039 (0.013)**
*pre*LAS Gain Score	0.115 (0.104)	-0.031 (0.027)
Sp. K Working Memory X *pre*LAS Gain Score	0.002 (0.006)	0.001 (0.002)
Fall K Cognitive Flexibility	0.537 (0.157)**	0.009 (0.061)
Fall K Attentional Focus	1.516 (0.672)*	0.548 (0.195)**
Fall K Inhibitory Control	0.201 (0.849)	-0.479 (0.211)*
Fall K Full-Day	2.784 (1.460)	-0.255 (0.588)
Parental Pre-K Care	-0.092 (1.721)	-0.414 (0.658)
Age of Kindergarten Entry	0.081 (0.180)	-0.156 (0.053)**
Not First-Time Kindergartener	4.270 (2.814)	-2.318 (0.98)*
Socioeconomic Status	2.106 (1.218)	-0.832 (0.413)*
Non-Hispanic, Non-Asian	0.571 (2.967)	0.419 (1.941)
Asian	5.748 (3.141)	1.926 (1.008)
Female	-0.479 (1.221)	-1.201 (0.417)**
Fall K Mathematics IRT	0.525 (0.097)***	0.031 (0.035)
EBRS Model		
(Intercept)	57.717 (1.625)***	14.654 (0.668)***
Sp. K. Working Memory	0.227 (0.051)***	-0.038 (0.012)**
EBRS Gain Score	0.576 (0.158)***	-0.104 (0.042)*
Sp. K Working memory X EBRS Gain Score	-0.005 (0.007)	0.002 (0.003)
Fall K Cognitive Flexibility	0.486 (0.164)**	0.018 (0.061)
Fall K Attentional Focus	1.581 (0.653)*	0.534 (0.211)*
Fall K Inhibitory Control	0.060 (0.845)	-0.458 (0.209)*
Fall K Full-Day	1.325 (1.557)	0.025 (0.621)
Parental Pre-K Care	-0.292 (1.539)	-0.383 (0.621)
Age of Kindergarten Entry	0.125 (0.165)	-0.165 (0.055)**
Not First-Time Kindergartener	5.211 (2.61)*	-2.485 (0.929)**
Socioeconomic Status	2.564 (1.165)*	-0.892 (0.404)*
Non-Hispanic, Non-Asian	0.271 (3.065)	0.485 (1.911)
Asian	6.316 (3.031)*	1.790 (0.969)
Female	-1.024 (1.204)	-1.067 (0.430)*
Fall K Mathematics IRT	0.620 (0.097)***	0.019 (0.034)

Note. All estimates centered at the ELL group mean of each continuous variable. **p* < .05, ** *p* < .01, ****p* < .001

Table 18
Conditional Latent-Basis Growth Curve Parameters

Parameter	*pre*LAS Model Estimate (*SE*)	EBRS Model Estimate (*SE*)
Latent Intercept Conditional Intercept	56.12 (1.34)***	57.72 (1.63)***
Latent Intercept Residual Variance	92.23	84.78
Latent Slope Conditional Intercept	14.97 (0.62)***	14.65 (0.67)***
Latent Slope Residual Variance	6.85	6.39
Intercept-Slope Covariance	-5.25 (3.56)^	-3.56 (3.17)^
Latent Slope Factor Loadings		
Sp. Gr. 1	Fixed at 0	Fixed at 0
Sp. Gr. 2	1.38 (0.05)***	1.38 (0.05)***
Sp. Gr. 3	2.42 (0.03)***	2.41 (0.03)***
Sp. Gr. 4	Fixed at 3	Fixed at 3
Residual Variances		
Sp. Gr. 1	27.26	29.53
Sp. Gr. 2	52.53	51.86
Sp. Gr. 3	37.36	37.29
Sp. Gr. 4	22.90	22.96
Fit Indices		
R^2 Latent Intercept	.47	.51
R^2 Latent Slope	.16	.19
RMSEA	0.053	0.056
SRMR	0.033	0.033
BIC	8943.506	8926.687
χ^2 (*df*)	60.649 (33)	63.989 (33)

Note. *p*-values not provided for fit indices and residual variances. ^$p \geq .05$ ***$p < .001$

*pre*LAS gains did not emerge as a positive predictor of later growth level or slope in mathematics. In the EBRS model, the growth model intercept regression on EBRS change revealed a significant and positive relationship, indicating that students' growth in basic English print knowledge skills predicts higher mathematics performance at the end of first grade. However, they see diminishing returns in this advantage through fourth grade due to the negative effect on the growth slope. Nevertheless, the over half-point advantage at the end of first grade translates into only a quarter-point advantage by the end of fourth grade due to the negative slope coefficient (i.e., -.104 * 3). Given the likely relationship between basic reading skills and receptive/expressive vocabulary in predicting mathematics performance (e.g., Chow & Ekholm, 2019), a third model was run with both *pre*LAS and EBRS change included in the same model to check the robustness of the EBRS finding. This model did not change the pattern of results for the EBRS or *pre*LAS effects on the growth model intercept and slope, indicating that kindergarten gains in basic reading skills in English are a unique predictor of mathematics performance level and growth beyond development of English expressive and receptive vocabulary. In none of the models did working memory significantly moderate the effect of *pre*LAS or EBRS gains on mathematics performance level or growth.

The effect of working memory in this model also differed slightly than what was identified in Study 1. This difference is largely driven by the difference in the model estimation technique and the different sources of data. Because Spring of first grade rather than kindergarten was used as the intercept, the developmental relationship between students' intercept and subsequent slope varies between studies. Increases in working memory predict higher performance level but given the subsequent pattern of development after first grade, this higher performance level does not necessarily translate into compounding growth over time as it does in kindergarten when constraining the effect of working memory to be equal across language groups as in Study 1. When left unconstrained across groups, working memory shows a relatively similar though non-significant effect on the slope among ELLs in study one as observed here in study two. However, Study 1 showed that the model is better fit with working memory constrained across groups, likely because there is far more power among EPSs to estimate the effect, thus greater efficiency in model estimation is gained when constraining the effect across groups. The results here, however, show that for ELLs, the first-grade advantage gained from being higher in working memory at the end of kindergarten has slightly diminishing returns through the next three years. Scoring one point higher in working memory predicts an increase in .226 of a point in mathematics at the end of first grade, though this is offset by a slightly flatter slope (working memory coefficients for slope of -0.039 and -0.038 in *pre*LAS and EBRS models, respectively). By the end of fourth grade, this negative slope coefficient translates into the advantage of working memory from first grade mathematics being cut in half. Another important aspect to note about the conditional model is that the BIC value is significantly lower than the unconditional model, and it is significantly lower than any of the mixture models with two or three classes. This may suggest that including the baseline covariates may be more effective for building a better-fitting model (relative to the unconditional model) than modeling latent mixture classes, particularly in terms of balancing model complexity with parsimony (Kaplan, 2002; 2009; 2014).

Study 2 Discussion

In Study 2, I investigated the presence of mixture classes of mathematics growth from Spring of first to Spring of fourth grade among 303 initially low ELP ELLs. I predicted that multiple mixture classes of growth would be extracted, helping further identify students who were potentially identifiable as at-risk learners between grades 1 and 4. Upon establishing the presence of qualitatively distinct growth patterns, I planned to implement a regime-switching model to detect stability in mixture class membership and predictors of that stability. However, no qualitatively distinct mixture classes could be extracted from the distribution of latent growth curves. Instead, ELLs' mathematics growth trend from first to fourth grade seemed to be best characterized as a single, highly variable pattern of growth. Though this does not align with the initial prediction, this finding is nonetheless informative of the pattern of ELLs' mathematics development. Specifically, the population of lower ELP ELLs does not contain subpopulations that change in systematically heterogenous ways. Rather, the significant majority of ELLs grow variably within the context of a single, normative pattern. Variabilities that occur beyond this normative trajectory can be conceptualized as disturbances from an average pattern as opposed to distinct classes of growth. Thus, teachers must utilize qualitative and quantitative data to monitor low ELP ELLs' response to instruction and intervention and cross-validate this information with national, state, and local norms (Albers & Martinez, 2015).

However, these results do not suggest that students with mathematics learning challenges or at-risk learning characteristics do not exhibit their own distinct set of characteristics. Rather, these characteristics do not manifest in qualitatively distinct development patterns relative to other low ELP ELLs when viewing mathematics skills as a composite form of various subskills. Other studies have established latent subpopulations of general mathematics competence growth among Latino students in kindergarten and first grade (Hong & You, 2012). In early elementary samples, Jordan et al. (2006, 2007) extracted latent subpopulations of mathematics growth using a more granular measure of mathematics that specifically tapped early numeracy ability. The common characteristic of these studies was that both empirical and substantive assessments of the mixtures provided a basis for the distinction among mixture classes. In the current study, the mixture procedure extracted a very small subpopulation with model indices showed very strong class separation (entropy > .90) and marginally improved relative model fit. However, the qualitative importance of the very small mixture class was minimal given that the growth curves differed meaningfully only in the intercept, one factor loading for the slope in grade one, and the residual variances. Various forms of academic performance data for these students must be leveraged to capture the factors that contribute to learning difficulties in addition to the behavioral, social-emotional, and cultural characteristics that drive learning, as it may not be the developmental trajectory itself that establishes different subpopulations of students. For example, given the differential predictive capability of early mathematics subskills (e.g., Nguyen et al., 2016), further research ought to attend to domain-specific mathematics subskills, as there is practical and theoretical justification to believe there may be differential profiles of development in specific skill areas among ELLs considering their consistent lower performance compared to EPSs.

These results provide additional potential insights into the learning process itself. For example, high variability in development may indicate difficulty building on mathematical knowledge to apply to new concepts and procedures, especially as fraction and word problem skills become more heavily recruited in intermediate elementary. The decreasing magnitude of the mathematics measure correlations from first to fourth grade suggest deteriorating stability in individual differences over time, which is then captured in the high and varying residual variance of the growth model. Students who are already higher performing in this area may be more buffered through these transitions with stronger mathematics prior knowledge and cognitive abilities (as they relate to mathematics) because they may be more capable of effectively reasoning through the new content and language demands. Further investigations into predictors of performance across these transitions is needed to understand learner characteristics that facilitate acquisition of fraction and prealgebraic skills among ELLs. Fractions in particular have been a strong focus of intervention and developmental research in recent years given their critical role in continuing to perform in mathematics (Fuchs, Schumacher, Sterba, et al., 2014; Resnick et al., 2016; Rinne et al., 2017). Fractions are also difficult because the language utilized to express their mathematical may be unique compared to other areas of mathematics learning, potentially placing a new academic language demand on ELLs, especially those who are less English proficient (e.g., “part of”, “percent”). Such students require additional instructional scaffolding and considerations for prevention and intervention if they are unresponsive to general classroom instruction and language (L1 and L2) supports.

Following the identification of a single class of growth, I investigated the extent to which growth in ELP across kindergarten and the interaction between ELP growth and working memory was predictive of later growth trends in mathematics. The absence of a significant effect of *pre*LAS change on later mathematics growth indicates that the rate at which students gain oral ELP is not a meaningful predictor of later growth in mathematics net of other demographic and cognitive characteristics. However, the significant positive effect of change in EBRS across kindergarten on growth intercept and significant negative effect on slope is suggestive of a potentially stronger relationship between domain-specific performance in English language reading areas, such as phonemic awareness and print knowledge, rather than general expressive and receptive vocabulary. Greenfader (2017) found a similar relationship between Hispanic students’ kindergarten-entry EBRS performance level and mathematics performance at the end of second grade. Quickly acquiring basic print skills likely helps students build early mathematics skills by improving the extent to which students can reason with print-based mathematical information. One limitation to this change score approach is that students who changed more were likely to be lower scoring at kindergarten entry to begin with, partly because of the scaling of the measure. Another potential explanation is that students with lower basic English print knowledge might grow more because they receive more instructional supports or interventions early in kindergarten, which may accrue and transfer to mathematical knowledge through elementary school. Students who are already high performing on EBRS may receive fewer of these supports, and their skills may not build at the same rate as students who received more intensive supports. In fact, many students ($N = 30$) actually decreased in their EBRS score from Fall to Spring, and 10 students did not change at all. With the exception of the already high scoring students, students who grew more potentially responded to literacy instruction and ELL supports more robustly, and this response may have help them acquire the strategies and skills to perform higher in mathematics. Another alternative explanation is that students who changed more were inherently more likely to respond more strongly to instruction, which means they were able to accrue academic content knowledge and skills in all subjects. This hypothesis requires more investigation into bivariate processes of academic growth among ELLs to examine the processes that simultaneously compound and predict with each other.

While both print knowledge and oral language are important for educational success among ELLs and EPSs alike, the current results seem to point towards more basic academic print knowledge growth (EBRS) as a stronger direct route to mathematics knowledge (over and above other skills) among ELLs. As such, greater growth in oral language on measures like *pre*LAS score may not be an indicator of strong responders to instruction over and above other skills. Skill gains on the *pre*LAS may be more socially embedded and may not transfer as directly to mathematics skills a year later over and above other abilities (e.g., working memory, attentional focus) that are also crucial to curriculum access and instructional response. Chow and Ekholm (2019) found that language measures of syntax were more predictive of mathematics performance than receptive vocabulary and morphology among first and second graders. This may offer an explanation of the current finding, given that basic reading skills may map more onto the structure of language (i.e., syntax) as opposed to measures of vocabulary especially as advance into more complicated syntactical structures in the years following kindergarten (i.e., among the age group Chow and Ekholm [2019] studied). As Chow and Jacobs (2016) and Chow and Ekholm (2019) note, receptive vocabulary measures are some of the most frequently used measures of language in many mathematics studies. Although English vocabulary among ELLs is highly important in developing ELP, the current findings, as well as Chow and Ekholm's (2019) results, point towards more structural components of language as key predictors of mathematics performance, regardless of native language, though it may be particularly important as a gateway to accessing mathematics instruction for ELLs. Chow and Ekholm's (2019) results also controlled for ELL status, providing further evidence that the role of language syntax is unique of native language status (but not necessarily L1 or L2 proficiency).

Synthesizing the findings from Study 2 shows not only that ELLs change in similar ways when assessed in English between grades 1 and 4 but also that early gains in oral ELP do not explain significant variation in growth patterns. Rather, consistent with Study 1, working memory and kindergarten-entry mathematics performance emerged as significant predictors of grade 1 Spring mathematics performance level. Additionally, in investigating English basic reading skills as an alternative measure of English proficiency and reading, students' gains on this measure emerged as a significant positive predictor of grade 1 performance level but negatively predicted growth. Together, these findings show that the hypothesized patterns and predictors of growth among low ELP ELLs were not supported, and the theoretical mechanisms explaining change remain inconclusive.

However, these results are likely contingent on the methods that were selected for investigation. Specifically, alternative methods to identifying heterogeneities in growth should be explored in future research. Additionally, this study was primarily concerned with the extent to gains in one domain (ELP) predicted gains in another and what may moderate those gains. Because neither oral ELP gains neither predicted nor were moderated by working memory in predicting later growth, more research is needed to examine the extent to which the rate at which students acquire ELP is related to mathematics growth, if at all. At a theoretical level, it is important to investigate the ways in which language gains relate to development in mathematics given the other data suggesting that levels of ELP (and language in general) are highly related. From a practice and policy perspective, summative ELP assessments such as ACCESS directly assess students' development along a continuum of ELP, and schools typically attend to the rate at which students acquire ELP as an indicator of success of their language and general classroom instruction. Such large investments in measuring student growth ought to undergo rigorous quantitative evaluation to determine the use of such measures in instructional planning beyond English supports and reading. However, Chen and Chaloub-Deville (2016) also note in their work using the ECLS-K: 1999 that English reading ability may be the more critical language-based component of mathematics rather than oral English performance, as English reading explained all of the mathematics performance gaps between non-ELLs, former ELLs, and current ELLs. For this reason, combined with more recent evidence from Chow and Ekholm (2019), more nuanced research into the components of language that matter for mathematics learning among all students—but especially ELLs—must be conducted.

Last, the circumstances under which these measures exhibit the highest utility in predicting performance in other domains (i.e., moderation) is crucial to consider given the role that cognitive characteristics play in mathematics development and response to mathematics intervention (Fuchs, Schumacher, Sterba et al., 2014). Though the current study did not identify moderators of ELP in predicting mathematics, future research in mathematics and ELLs should consider the domain-general components of learning that uniquely and differentially contribute to trends in mathematics development. Examining conditional and differential processes (i.e., moderation and moderated mediation) is particularly important given the prior empirical work supporting the hypothesis that bilingualism can benefit executive functioning. As such, from an RTI perspective, it is imperative to examine the circumstances under which and for whom second language and executive functioning interact in predicting mathematics outcomes. While this extends largely from a developmental perspective, testing conditional and differential processes such as these can directly inform the skills (both domain-general and -specific) interventions target and when those skills ought to be targeted.

Limitations

The principal limitation to this study is the form of measurement of ELP. Although I tested multiple forms of potential English proficiency measures (EBRS and *pre*LAS), these are highly limited in scope compared to the measures used in schools. For example, the ACCESS 2.0 (World-Class Instructional Design and Assessment, 2014) is a comprehensive measure of English language proficiency that assesses domain-specific content language, oral language, and English writing ability. As such, the instructional and programmatical decisions that are made in schools that use such assessments are drawing on data that is much deeper and broader in scope. Growth in ELP is then based on a much deeper understanding of the skills that comprise ELP and a better statistical model of growth since the assessment is designed to assess such change. Another significant related limitation is that not all students were assessed in the same time intervals. While it is questionable how much this significantly impacts the current estimates (and extraction of mixture classes), this is a significant limitation of the ELP gain scores, as gains for one student may not be measured across the same time metric as another. This is likely significantly more impactful for the change scores as opposed to the growth model because change was assessed from Fall to Spring, introducing more variability in the time intervals of assessment rather than Spring-to-Spring measurements of mathematics. Additionally, though the definition of ELL used in this study was made conscientiously to better approximate the definition of language proficiency schools might use (in terms of the range of ELP captured on more comprehensive measures), alternative definitions should be considered when using the ECLS-K:2011 data and other datasets that employ less comprehensive measures of ELP.

A second significant limitation of this study is the timing of the measures used to assess the moderating effect of working memory. Specifically, ELP growth was assessed from Fall to Spring of kindergarten while working memory was assessed in Spring. As a result, it is possible that working memory is better characterized as a mediator rather than a moderator. This was a limitation of the available data since many students did not score high enough and were not old enough to receive a scale score in the Fall of kindergarten. Alternative scores for working memory were available for students who scored too low to acquire a scale score; however, the scaling of these alternative scoring methods differed between Spanish and English versions of the assessment. The minimum *w*-ability score for the Spanish WJ-III NR in Spanish was lower than the English version across the entire scale, so different scores corresponded to the same level of performance (Tourangeau et al., 2015a). Thus, this score was deemed inappropriate for the current analyses. The minimum standard score for the Spanish WJ-III was also lower than the English version; however, no students who took the Spanish version in the Spring of kindergarten obtained a minimum score lower than the minimum score in English. As such, this measure appeared less problematic in terms of the scaling properties. Psychometric properties notwithstanding, the timing of each measurement is important in terms of the temporal relationships between constructs.

Chapter V: General Discussion

Summary

The current studies addressed ELL-EPS differences in mathematics change patterns, potential heterogeneities in ELL mathematics development, and the interactions between working memory and language proficiency. Findings from Study 1 show groups exhibited consistent growth in mathematics with varying degrees of fluctuation in the extent to which autoregressive relationships account for the amount of change that occurs. Specifically, both ELLs and EPSs grew significantly through first grade then decelerated in grades three and four, though ELLs ended kindergarten nearly a standard deviation lower than their EPS peers. Considering the large gap at kindergarten entry, holding both ELLs and EPSs at the mean of Fall kindergarten mathematics closed essentially all of the achievement gap over time.

Results from Study 2 are suggestive of a different pattern in ELL growth once students are assessed in English. Growth from first to fourth grade was relatively homogenous in terms of the distribution of individual growth curves. While there are significant time-varying disturbances from the general trend, a latent-basis, nonlinear (decelerating) trend seemed to adequately capture the growth in English-based mathematics assessment performance among the subpopulation of low ELP ELLs under investigation. Nevertheless, student growth was highly variable while simultaneously being captured by a single underlying trajectory. After identifying a single-class growth model among ELLs, kindergarten year gains in *pre*LAS scores did not meaningfully predict growth from first to fourth grade, though there is evidence to suggest that gains in basic reading skills predict later mathematics performance level. Yet gains in basic English reading skills (e.g., phonemic awareness, letter-sound correspondence) positively predict higher mathematics performance at the end of grade one while they also predict a declaration in growth, essentially offsetting the initially intercept prediction. Furthermore, contrary to the initial prediction, working memory did not moderate the extent to which language proficiency gains predicted mathematics growth.

Taking both studies together, ELLs exhibit similar patterns of growth to their English proficient peers, though they end kindergarten performing significantly lower than EPSs. These gaps persist over time, though baseline mathematics, basic reading, and working memory help explain these differences. At the same time, key executive functioning skills, specifically working memory, similarly contributed to predicting development across groups. Among ELLs specifically, gains in English expressive and receptive vocabulary did not meaningfully predict patterns of development in English-based mathematics skills, though basic reading was strongly related to initial level and trend in later mathematics growth.

Support for Hypotheses and Research Questions

Although the majority of the original hypotheses were not supported using the current data, these findings nevertheless contribute to the meager base of methodologically rigorous studies on mathematics development among ELLs, specifically regarding pattern and distribution of change as well as domain-general predictors of growth (i.e., ELP, working memory). These results corroborate the well-established role of working memory in predicting individual differences in mathematics development while doing so among a particularly understudied subgroup of low ELP ELLs. Significantly more research needs to be conducted to understand how and why this relationship holds (i.e., mediators), yet these results may help guide the formulation of theory in targeting common elements of intervention across language groups. Additionally, results from this work suggest that intraindividual change occurs somewhat similarly among ELLs and EPSs. On the other hand, controlling for baseline characteristics shows that, beyond the additive growth across years, change among ELLs is not contingent on prior performance levels until third grade while prior levels predict subsequent change across all years for EPSs. However, this does not translate into vastly different trajectories, and these results do not align with the hypothesis that ELLs would exhibit greater plateauing relative to prior growth compared to EPSs. In fact, the results suggest the opposite: EPSs plateau more noticeably compared to ELLs. As mentioned in the Study 1 discussion, this may be due in part to the level at which students are performing in mathematics: EPSs, on average, are assessed in challenging content and thus reach a ceiling in their ability to make equal gains over time, thus generating the decelerating effect observed in the MG-LCSMs.

When examining distributions of growth among ELLs once they are assessed in English (i.e., first through fourth grade), there is a high enough level of homogeneity in the non-linear growth pattern to suggest a single, average trajectory best characterizes intraindividual change (though there are significant individual differences in terms of the deviations from that curve). The lack of cogent evidence for systematic heterogeneity in change is possible indicative of a greater level of developmental continuity among this subpopulation of ELLs than might be expected among EPSs, in part because EPSs constitute a much larger portion of students in the U.S. and thus contain far more potential subpopulations. Latent subclasses of growth curves can provide extremely valuable information on how and for whom interventions should be targeted and what schools can expect of student growth under business-as-usual conditions. The current results suggest that individual differences in developmental trajectories are highly variable though similar on average among ELLs while the pattern of intra-individual change is somewhat similar regardless of native language.

Implications and Future Directions

The nature of the gains captured in the current study is an important consideration for multiple areas of theory, research, and practice. First, these change patterns underscore the imperative for equitable access to instruction among ELLs (Robinson-Cimpian et al., 2016). Indeed, many adjustments to instruction need to be made to accommodate the learning of ELLs; however, many of these components are likely beneficial to all students. For example, opportunities to verbalize problem solving and thinking is a frequent recommendation in the explicit instruction literature to improve mathematics performance (Jayanthi, Gersten, & Baker, 2008). Doabler, Clarke, Kosty, et al. (2016) demonstrated this main effect of verbalization of thinking and problem solving across language groups among kindergarteners in a large scale-up trial of a tier 1 mathematics curriculum. Although there were not differentially positive effects for ELLs, all students increased their mathematics performance over kindergarten. This indicates the critical role of employing linguistically rich mathematics instruction, though further scaffolds are needed to support ELLs in closing early performance and access gaps. Consistent with developmental research (Vukovic & Lesaux, 2013), language accounts for mathematics knowledge development similarly across ELLs and EPSs. Similarly, other findings from the RTI literature indicate that language ability (across language groups) is a strong predictor of response to mathematics intervention in second grade (Powell et al., 2017). This is an important consideration for screening for students at risk for mathematics difficulties as well. Looking solely at mathematics skills, the expectations of “at risk” among EPSs may not reflect the same dimensions of risk among ELLs given the nature of early disparities that endure across early and intermediate elementary school. As such, the factors that predict mathematics development and adequate response to instruction need to be examined more carefully across and within grade levels.

These between-group growth patterns provide multiple insights into how practitioners and researchers must consider the role of RTI in mathematics development. While screening for the same skills across all students constitutes the basis of universal screening procedures, additional considerations for defining academic risk among ELLs, such as qualitative and quantitative indicators of performance via teacher-guided formative assessment, local norms/cut scores, and ELP data ought to be considered in practice throughout intervention progress monitoring and triannual screening procedures (Albers & Martinez, 2015; Hall, Moore, McMackin, Markham, & Albers, 2019). Such data, corroborated with other indicators of performance, can provide further insight into students' profiles of domain-specific skills. Using their knowledge and training in normative development and evidence-based assessment practices for identifying at-risk students, school psychologists can play a crucial role in leveraging this data within school-based problem-solving teams to inform data-based decision making (Albers et al., 2013).

Linguistic, literacy, and mathematics factors contribute to academic risk in mathematics, though the manner by which these factors are related to mathematics may differ for ELLs as their prior knowledge differentially accumulates over time. The mechanisms of these change patterns (and the implicit patterns of instructional response) requires further investigation, building on both the current study and prior work investigating longitudinal patterns of language-mathematics relationships (Chen & Chaloub-Deville, 2016; Hartano et al., 2018; LeFevre et al., 2010; Vukovic & Lesaux, 2013). Additionally, more research in this area pertaining to the mechanisms of building mathematical knowledge is required. For example, the relationship between how procedural and conceptual knowledge is built among lower ELP ELLs may help unpack the driving forces of developmental change, especially with regards to students' progressing through mathematics domain competencies, such as those assessed in the ECLS-K measure. Identifying these mathematical learning mechanisms can more effectively inform the targets for change in both universal, secondary, and tertiary instructional contexts. While this study does not provide a causal basis for low ELP ELL and EPS differences over time, it is nonetheless clear that significant research must be devoted to understanding interventions to target early mathematics gaps and the resulting developmental patterns as well as how these performance gaps can be mitigated over time.

Such research requires both analytical savviness and strong theoretical bases for how such interventions can sustain over time (Bailey, Duncan, Odgers, & Yu, 2017). Alternative modeling techniques can provide more empirical basis for how mathematical knowledge is built and sustained. For example, a state-trait framework may provide more insight into the breakdown in stable individual differences as opposed to autoregressive patterns in mathematics development (Bailey, Watts, Littlefield, & Geary, 2014). Recent work has shown the value of state-trait model in evaluating intervention effects by decomposing intervention effects into impacts on stable performance across grades versus what helps students continue to perform between grades (Watts et al., 2017). Such modeling could provide a stronger basis for stable time versus time-specific variability. Although such work operates within the larger theoretical perspectives of development and education, it is exactly that which can help better inform the selection of measures, frequency of measures, and the development and targeting of feasible, efficient, and malleable intervention mechanisms. These theoretical and developmental approaches to examining mathematics achievement among linguistically diverse individuals can aid in understanding how to best leverage RTI data, instructional practices, and policy to close performance and access gaps in early schooling and sustain mathematical knowledge over time. The current results suggest that, even when considering the effect of prior mathematics achievement on growth, the fundamental pattern of change between ELLs and EPSs is not all that different. Causal intervention and developmental evidence that can establish mechanisms that drive intraindividual change among ELLs and EPSs will help identify the malleable factors across groups that can be capitalized to efficiently and equitably accelerate mathematics growth among ELLs with multiple levels and contexts of instruction.

Implications for Theory

While the specific hypotheses were not supported, the results from both studies nevertheless confirm some of proposed process identified in Figure 1, Chapter 1. Indeed, language proficiency as a category (i.e., Study 1) showed that there is strong evidence for significant differences in initial growth levels that sustain across elementary school. Study 2 showed that among those identified as low ELP ELLs, expressive and receptive vocabulary gains do not uniquely predict mathematics growth in English yet gains in basic English reading skills do. Though these skills represent early literacy abilities rather than the broad construct of ELP, these skills are nonetheless likely important factors for students' capacity to respond to general classroom instruction and language-specific services (e.g., dual-language instruction). Across both studies, working memory emerged as a significant predictor of mathematics growth, albeit in different ways. Study 1 showed that working memory predicted higher levels of mathematics performance at the end of kindergarten and a compounding effect on growth such that higher working memory performance predicted greater accrual of mathematics skills over time. On the other hand, working memory positively predicted higher mathematics performance at the end of grade one but simultaneously predicted a decline in growth, similar to EBRSs. Considering the model techniques were quite different between Study 1 and Study 2, the general takeaway is that early working memory tends to predict an increase in both kindergarten performance and gains for both ELLs and EPSs, while solely among ELLs, those same increases in working memory translate into higher performance levels but greater decline in subsequent mathematics growth when predicting first grade performance level and subsequent growth.

One potential explanation for this is that early higher performance as a result of higher working memory may help students accrue more mathematics skills and sustain increases in growth over time. On the other hand, when predicting first grade performance levels and later growth from kindergarten working memory, the effect of being higher in first grade could relate to less capacity to sustain greater growth over time, possibly due to the increasing complexity of mathematics performance. Indeed, when leaving all regressions unconstrained across groups, the effect of working memory on mathematics slope among ELLs was non-significant with a point estimate close to zero (though the BIC suggested unconstrained working memory did not improve fit when all other regressions were constrained to equality across groups). This may reflect the similar pattern observed in Study 2, though because the Study 1 model was significantly different both in its structure and the waves of data used, the effect of working memory emerged as close to zero rather than negative and significant as the slope estimate is estimated starting at kindergarten when more growth among ELLs occur. Additionally, the estimate of working memory in Study 2 reflects an effect on students' performance level and growth solely in English (except for the nine students at the end of first grade assessed in Spanish). Relationships of working memory to earlier performance and growth in Study 1 may reflect an ability of working memory to offset some later declines as students tend to grow more across kindergarten than in later periods, as such the decline predicted in the Study 2 model is washed-out in the Study 1 model. In Study 2, advantages of early working memory for first grade performance are diminished as students age because there is greater likelihood of deceleration in mathematics growth between first and fourth grade. Such relationships can be observed most directly in the intercept-slope covariances of each model. In Study 1, the unconditional model covariance was constrained across groups but was negative and significant, indicating that performing higher in kindergarten was associated with flatter constant growth through fourth grade. Including the covariates reversed the sign of this covariance, suggesting that after accounting for the variance the covariates explained in intercept-slope variances/covariances, there is a significant positive relationship between performance level and constant growth. Because the covariates contribute to reversing this sign, it is then intuitive that working memory would positively predict both performance level and constant change in mathematics. In Study 2, the intercept-slope covariance of the unconditional growth model was negative and nonsignificant (students who performed higher potentially grew less). Including the set of covariates explained a significant portion of this covariance, leaving it still nonsignificant and negative. As such, this negative sign is then consistent with the effect of working memory, that because working memory is in part explaining the negative intercept-slope covariance, it is intuitive that as one of the stronger predictors of performance level and growth, it would positively and negatively predict level and growth, respectively.

These relationships between prior and concurrent level and subsequent trend are indicative of how skill and performance compound over time. This process reflects an unobserved cascade of effects, as prior status holds some relationship to the developmental pattern that ensues. Identifying the specific cause of that relationship is untenable in this work;

however, it is important to theoretically consider why it occurs and the potential counterfactual of its absence. Developmental systems and cascades (Ford & Lerner, 1992; Lerner et al., 2011; Masten & Cicchetti, 2010) posit that multiple dimensions of change influence the change of each other, and that the context in which those domains develop influences the strengths and patterns of cascading relationships. This is specifically the process by which ELP and working memory can be hypothesized to relate to mathematics development. Similarly, this is implicitly the process by which change is theorized to occur in Figure 1. The results here indicate that entry-level ELP is related to sustained lower levels of achievement. Again, a causal inference about this effect is untenable in the current work; however, because the set of baseline covariates explains all of the between-group differences, level of ELP and factors associated with ELP level explain these sustained gaps. Based on this, there are seemingly many processes that are simultaneously related that appear to in part generate these sustained gaps. However, Study 2 shows that growth in basic English reading skills – net of domain-general and -specific skills – relates to more proximal higher levels of mathematics performance while also predicting a distal fadeout through the end of fourth grade. This indicates that the rate at which students acquire early reading then cascades (whether directly or indirectly) into temporarily higher mathematics performance, though these higher levels do not compound over time. In essence, early English reading skills among ELLs cascade in two different ways in relation to mathematics. Expressive and receptive vocabulary, though highly important on its own, does not cascade in the hypothesized fashion in this context. As previously mentioned, this may be due to the content-specificity of the EBRS measure and its contextual relationship to mathematics: basic reading skills likely relate more directly to being able to access mathematics content. The current analyses approached this from essential a sequential process model (prior changes predicting future changes). Bivariate process techniques can assess cascade theories more effectively, though careful attention must be paid to distinguishing among between- and within-individual temporal differences in cross-lagged processes (Berry & Willoughby, 2017).

The current findings on working memory can be framed in terms of cognitive load theory (de Jong, 2009; Sweller, 1988; Sweller et al., 2011). Figure 1 theorizes a unique relationship of cognitive attributes to gaining mathematics skills as well as a simultaneous relationship with English proficiency for this reason. ELLs are likely under greater cognitive load while learning compared to EPSs simply due to the added language barrier. Assumedly, this varies significantly based on the instructional context and the strategies of the general classroom and language instructional program instruction. Nevertheless, the dynamics of learning likely vary due in part to ELP. Working memory plays a unique role in mathematics development, and it is also a core component of cognitive load. As such, its relationship to mathematics among ELLs may provide highly important insight into how cognitive load relates to mathematics. However, findings here suggest that working memory is not differentially predictive of mathematics development across language groups. One perspective is that I controlled for a significant majority of the observable differences among ELLs and EPSs, so the remaining differences are highly limited, leaving working memory with a highly similar, unique effect across groups. The other side of this explanation is that once I controlled for these between-group differences, there are very few meaningful between-group differences in mathematics development such that the role of cognitive load matching on all of these other observable differences then functions essentially identically in each group. Findings from Study 2 show a slightly different effect of working memory specifically among ELLs (though this should not be confounded with the fact that the modeling process was significantly different in Study 2). Nevertheless, Study 2 showed working memory continued to evidence a significant unique relationship to later mathematics development. The lack of an interaction with language growth, though, indicates that regardless of how ELP is measured (*pre*LAS or EBRS), working memory does not affect the strength of language proficiency in predicting mathematics growth. As such, the relationship between working memory and ELP based on Figure 1 must be further refined to reflect the uniqueness of working memory in particular.

To that end, both working memory and language processes need to be tested with more refined connections between theory, methodology, and measurement. Unique working memory domains ought to be tested as both cross-sectional and longitudinal (e.g., cross-lagged models) predictors of mathematics development, as these may have different relationships with how ELP may relate to cognitive load. Verbal working memory clearly plays an important role in the development of mathematics; however, given its emphasis on verbal storage of numbers, it may inherently contain more shared variance in mathematics ability than measures of visuospatial working memory or other verbal listening recall tasks. Additionally, controlling for the covariates in these analyses essentially removed all observable differences between ELLs and EPSs leaving little for working memory to differentially predict, but that does not necessarily imply that potential mediators between working memory and mathematics development operate with the same strength across groups. Moreover, the relationship between ELP and mathematics must be assessed with greater detail to detect which components of ELP are most impactful in facilitating mathematics development, and equally importantly, how L1 relates to both of these development processes. ELP frequently is used as a catchall term; however, the ELP assessments used in schools are in fact quite diverse, assessing students in oral language, writing, and reading skills. The course of cascades between ELP, instructional response, and mathematics performance thus needs to be more scrutinized to identify which, if any, unique paths between these areas exist.

Testing such refined theoretical models can help lay a groundwork for targeting intervention and instructional practices that are shared across ELLs and EPSs as well as strategies that are unique to each group, increasing the efficiency and effectiveness of empirically-supported intervention approaches such as explicit instruction (Jayanthi et al., 2008). Specifically, finding similar and unique ways for each group to reduce cognitive load and maximize the efficiency with which students can build, use, and retain mathematical knowledge. In turn, this can help offset potential differential cascades that may sustain long-run performance gaps across groups.

Implications for Research

The findings from the current research provide both confirmatory and new findings related to studying relationships between mathematics and language proficiency. Results mainly from Study 1 confirm evidence from prior work that ELLs and EPSs change in relatively similar ways across early and intermediate elementary school, though without controlling for key baseline covariates, significant achievement gaps throughout those years remain (Roberts & Bryant, 2011). While the current results confirm such prior findings from previous cohorts of ECLS-K data, the direct testing of how working memory operates across language groups (as well as the inclusion of a larger set of covariates) provides further insight into the domain-general skills that are associated with developmental change in mathematics. However, I extend the findings of Study 1 in Study 2 by investigating the distribution of growth in mathematics (assessed in English) for systematic subdistributions of growth curves (mixtures). Those analyses showed no substantive presence of mixtures, though I observed significant heterogeneity in growth curves.

However, because this heterogeneity appeared more reflective of a general characteristic of the population of interest, the modeling approach did not extract qualitatively distinct mixtures. The main implication of this is that change among ELLs occurs heterogeneously, though not in a seemingly systematic fashion so as to suggest subpopulations of “movers” or “stayers.” This is not to say that systematic forms of heterogeneity in growth do not exist at all among ELLs; mixture modeling is one formulation of capturing heterogeneity. If future research is concerned with investigating heterogeneity in mathematics performance among ELLs, alternate specifications ought to be considered. Indeed, robustness to alternate methods and replicability in developmental research is critical to generating knowledge of developmental processes (Duncan, Engel, & Claessens, & Dowsett, 2014). Nonetheless, GMM is particularly relevant to RTI research because it assumes subpopulations of growth curves are comprised of distributions and can better reflect differences in individual differences that can inform how to establish RTI tiers whereas many other methods assume homogeneity within each latent class. That said, mixture classes of growth curves did not seem to adequately capture heterogeneities in development. It is possible that there are wave-specific mixtures that could be extracted with other latent variable analyses such as latent profile analysis, though these heterogeneities do not manifest in differential developmental trajectories. Additionally, there are a number of other ways to detect heterogeneities in growth curves. Another method of examining heterogeneity in growth is latent class growth analysis (LCGA). Although highly similar to GMM, LCGM assumes homogeneity within classes (Kaplan, 2002). Extracting not only a class but a distribution of growth curves is one of the principle advantages to GMM (Kaplan, 2002).

However, assuming within-class homogeneity may have a strong theoretical basis, potentially making it a stronger candidate for examining heterogeneity among ELLs. Another example of examining heterogeneities in growth is the recently developed semtrees package in R (Brandmaier et al., 2013). This method, extending from classification and regression tree (CART) analysis, employs recursive partitioning based on a set of covariates to extract heterogeneities in structural relationships in a SEM while maximizing within-partition homogeneity. Recent work has compared growth mixture modeling to semtree analysis, finding that both methods are effective in separate ways in uncovering heterogeneities (Jaccobucci et al., 2017). As Jaccobucci et al. discuss, GMM can be based solely on heterogeneities in the outcome variable (i.e., the indicator variables comprising the latent growth factors), whereas semtrees extracts heterogeneities in the outcome (i.e., a growth model) solely based on a set of covariates. As aforementioned, a number of studies have demonstrated the substantive and empirical importance of mixture modeling in developmental and educational science. However, given the mixed empirical support for mixtures in the current study, alternative methods ought to be employed to explore heterogeneities in development among ELLs, provided there is a meaningful theoretical reason for doing so.

Methodological implications aside, the primary implication for future research from both Study 1 and Study 2 is that pattern of development in mathematics does not occur markedly different among ELLs compared to EPSs, even after controlling for key individual differences. What remains to be further investigated is the manner by which early mathematics and cognitive gaps occur, how they are sustained over time, and how policy and instructional interventions can promote academic success among ELLs. This is not based solely on the fact that, many times, ELLs are defined based on an achievement gap to begin with. While differences in language performance were used to establish ELLs in the current study, that distinction was not chosen assuming that would inherently translate to achievement gaps. Although this research showed performance gaps are most evident in domain-specific skills (i.e., mathematics), ELLs also performed nearly two-thirds of a standard deviation below EPSs in working memory at the end of kindergarten (controlling for Spanish administrations had no effect on this). As such, these early general and specific developmental process require significant further investigation to understand when and how these gaps arise. Indeed, much of these gaps are ecological and sociological in nature in that interactions between early development, experiences, and contexts are socially stratified relate to ELLs differently than EPSs, resulting in highly different cascades of early relationships prior to starting school. A number of significant sociological studies have been conducted to understand non-native English speaking and immigrant children's schooling experiences and early contexts. However, significant further research needs to identify malleable factors of change to promote and sustain successful academic, cognitive, and social-emotional-behavioral development in a culturally responsive manner. This relies on convergence of methods, theory, and substantive research hypotheses to empirically and qualitatively examine how early developmental cascades propagate throughout schooling and the risk/protective factors that affect those developmental systems.

Although the specification of ELLs in the current study was quite specific and may not be generalizable beyond studies using the *pre*LAS as a screener of ELP, there is significant variability in the definition of "ELL" in practice. The current study took a data-based approach to distinguishing among English proficient students that is ideally more reflective of the fact that ELL is not based only on parent-reported primary language, nor is it based only on a cut score with little empirical basis. Rather, the current study sought to establish the population of ELLs based on both parent-reported home language and the degree of separation of ELP between students who do and do not speak English as a primary home language, knowing that there is also a range of ELP among all students at the start of kindergarten regardless of primary home language. As such, an approach to defining ELLs that relied on native language and relative (instead of absolute) ELP is intended to generalize more to how schools observe and utilize information on ELP from both policy and instructional perspectives. Based on this specification of ELLs, both groups exhibit significant variability in individual growth curves; however, on average, models from Study 1 suggest intraindividual change does not occur meaningfully different. Any differences observed between those change patterns appears to be due to the significantly different sample sizes, resulting in very different levels of variance and parameter estimation precision. That said, the MG-LCSM model results suggest that, on average, EPSs exhibit slightly steeper gains from kindergarten to first grade. However, as aforementioned, this appears likely due to the fact that there is significantly more heterogeneity in the mathematics performance among EPSs. The distribution of performance among EPSs becomes increasingly left-tailed as students age, resulting in a larger decrease in change scores among EPSs and greater deceleration in mathematics growth over time. Controlling for baseline covariate results in closing the achievement gaps by the end of grade four, though the pattern of growth remained slightly different.

From Study 1, the finding that working memory does not operate differently across groups suggests that, net of other factors, working memory is uniquely associated with mathematics change regardless of kindergarten-entry mathematics performance and categorical English proficiency. However, based on results from Study 2, oral English proficiency gains do not predict level or trend in mathematics growth once assessed in English. The finding that basic

English reading performance predicts level (but not trend) implicates gains in early basic print knowledge as a key lever in building mathematics proficiency by the end of first grade, though does not appear significantly related to subsequent mathematics growth. To further unpack the role of working memory in the hypothesized developmental relationship between ELP gains and mathematics growth, I tested whether working memory moderated the relationship of ELP gains to mathematics building on the hypothesis that low working memory may impede mathematics development and that, as a result, ELP may become a stronger predictor of mathematics at lower WM levels as language proficiency gains may help offset lower working memory capacity. The lack of the hypothesized moderating notwithstanding, working memory continued to evidence a main effect on mathematics performance in terms of both level and growth trend.

Across both studies, working memory shows unique predictive capability among ELLs despite their linguistic (and thus in many ways cognitive) context of learning. Future research needs to corroborate this relationship of working memory in light of relationships with other WM subdomains. As aforementioned, backward digit span tasks – though frequently employed as a measure of central executive working memory – may be too strongly related to mathematics to offer the most effective estimate of the effect of working memory. Additionally, backward digit span tasks do not capture all aspects of working memory, and the extent to which it captures phonological loop versus central executive is debated in the literature. Gathercole et al. (2004) established backward digit span as a central executive measure of working memory as a result of its complex span structure: the authors established an age-invariant three-factor solution to working memory structure comprised of verbal-only span (e.g., digit span), complex-span (including backward digit span), and visual-spatial span (e.g., block recall). By this definition, BDS tasks such as WJ-III Numbers Reversed would be classified as central executive, though this formulation of central executive is not consistent across studies. For example, Swanson et al. (2018) used conceptual span, listening sentence span, and updating tasks to reflect the structure of the central executive component of working memory. Given the timing of the ECLS-K: 2011 study, the use of backward digit span is likely reflective of the understanding of WM based on research such as Gathercole et al. (2004). Indeed, this is based on some of the most popular formulations of WM, extending from Baddeley and Hitch (1974) and, with respect to central executive in particular, Baddeley (1996).

Such varying hypothesized structures of working memory is a catch-22. On one hand, a variety of measures to ostensibly measure the same latent construct of central executive working memory helps corroborate empirical hypotheses across different kinds of studies and serves as robustness checks against instrumentation. Conversely, this variability in measure selection fosters a debate about what constitutes central executive working memory, and these different formulations make it difficult to identify specific mechanisms of change that can be leveraged as malleable components of mathematics intervention. For example, the complex span tasks tapping the central executive in Swanson et al. (2018) are likely more complex tasks than the Gathercole et al. (2004) formulation of complex tasks, including backward digit span and may have different implications for moderating intervention effects (e.g., Fuchs, Schumacher, Sterba et al., 2014) or predicting mathematics development and intervention response (e.g., Powell et al., 2017). To that end, it would behoove future studies to clearly distinguish the rationale for the use of specific measures of working memory. Given the multidimensionality of working memory, the process through which distinct areas of working memory are hypothesized to predict outcomes or interact with concurrent predictors ought to be clearly elucidated. Doing so can significantly increase the specificity with which working memory and its subdimensions can be established key predictors, moderators, and/or mediators of mathematics outcomes, helping inform developmental and cross-sectional observational research as well as the design and testing of interventions and testing of treatment-by-aptitude heterogeneity.

Implications for Practice

The extension of these findings to practice is not direct, and it is important to clarify what aspects of this study can and cannot be extrapolated to practice. First, the two studies herein included very little information regarding the schools in the study, the classroom environments, or the types of instruction students were receiving. In that light, it is difficult to draw inferences from this work regarding the role of schools in student development. Such information is partly captured in the unexplained variance of the models, and estimates provided herein are corrected for the fact that students are clustered in specific schools and units of sampling. The information about schools that is included is minimal and is used to control for group and individual differences that may confound the relationship between initial mathematics performance and subsequent development as well as the relationship between WM and mathematics outcomes. For example, full- versus part-day kindergarten was included as a control given the numerous studies investigated how that difference may influence student academic and behavioral outcomes. Because there were no specific a priori hypotheses about the effects of, for example, pre-kindergarten care or full/half day kindergarten, further attention to those effects will not be given. Moreover, these results do not generalize to the practice of RTI itself (e.g., the specific screening procedures, the choice of interventions, and frequency of intervention progress monitoring). Instead, these results generalize to RTI in terms of how RTI builds upon knowledge of mathematics development and the predictors of that development. What these results can say about RTI, however, is four-fold.

First, even after controlling for a variety of baseline characteristics, ELLs and EPSs change in similar ways, albeit at different levels. Because achievement gaps appear relatively consistent across time, this would suggest that early interventions can, at the very least, potentially narrow achievement gaps relative to the gap size in kindergarten, even if the intervention effects fade out. Intervention fade-out has been the focus of a number of recent studies with Bailey, Duncan, Odgers, and Yu (2017) presenting cogent theoretical arguments for sustaining early childhood intervention effects. Some of Bailey et al.'s (2017) core arguments pertain to the timing of interventions ("foot in the door" techniques) and the role of "sustaining environments." For ELLs, both approaches are crucial in moving the needle on early mathematics achievement given the cooccurring development of both language and mathematics. Both factors are likely more malleable at earlier ages, which inherently is more preventative. However, the sustaining environments aspect is also highly important given that the targets of mathematics instruction shift over time and the language of mathematics increases in complexity. As such, sustaining the linguistically rigorous mathematics instruction across elementary school may help not only target more sensitive periods in development but also maintain the target skills and contexts that may offset poor academic trajectories.

Second, working memory appears to relate to mathematics development in similar ways across groups, suggesting that individual differences in working memory within each group operate similarly across groups. Coupled with the findings from Study 2, working memory relates to mathematics development similarly for all individuals in these studies, and does not moderate relationships between ELP gains and mathematics development. In essence, it serves as a similar risk/protective factor for mathematics development regardless of language level, and it does not create a heterogenous relationship between language gains and mathematics.

Third, the distribution of mathematics growth among ELLs is relatively uniform. Though there is significant, time-based heterogeneity in the extent to which this uniform growth curve captures individual development (i.e., residual variances change over time), there do not appear to be specific subdistributions of students. In an applied setting, this may imply that students could respond similarly to significantly different forms of instruction and that is the dosage of instruction on core, malleable skills that may be the key to improving performance. On the other hand, significantly different subdistributions of mathematics growth may indicate that certain classes of students have more unique needs (i.e., differential risk or differential high growth) that extend beyond what is captured by an average trajectory. The current findings suggest that although ELLs enter school with a variety of different experiences, backgrounds, and skills relative to EPSs, that does not necessarily imply that their needs for rigorous mathematics instruction is markedly different. Rather, as aforementioned, making sure all students have access to the same grade-level content (Robinson-Cimpian et al., 2016) seems to be the core component to building mathematics abilities among ELLs. Net of all other factors, differences in early mathematics achievement alone explain a significant amount of the EPS-ELL gaps through fourth grade. As such, screening procedures that can accurately and efficiently identify mathematics fluency deficits and brief, targeted interventions that can build early numeracy skills in the general education classroom as well as in tiered instruction may help mitigate some of these gaps over time. Again, while few results from this study can be directly applied to practitioners' work, these results do help provide further insight into the general mathematics trends that may be expected to occur in school settings and can help guide the manner by which state, district, and classroom level practices are coordinated to the developmental trajectories of ELLs and EPSs.

Fourth, there is mixed evidence about the role of the rate at which students acquire English proficiency skills in acquiring mathematics skills while controlling for the array of baseline measures, including prior mathematics performance. Gains in expressive and receptive vocabulary, typically viewed as key developmental skills given the popularity of them in ELL research (Swanson et al., 2018), did not predict mathematics outcomes a year later or the subsequent growth trends net of the other covariates. However, gains in English reading skills such as phonemic awareness, print knowledge, and sound blending significantly positively predicted mathematics performance at the end of first grade while simultaneously predicting a deceleration in mathematics growth. Given the content that the *pre*LAS and EBRS measures assessed, one explanation could be that the content of the EBRS measure is more indicative of students' access to core instruction and grade-level content. Though English expressive and receptive vocabulary is highly important in both social and academic contexts, it may be that basic reading skills translate into greater access to instructional content across subjects and may aid students more directly in navigating the print and phonemic knowledge required for accessing English mathematics content over and above prior mathematics knowledge. Simultaneously, however, steeper gains in EBRS may be an indicator of quicker or more effective response to school-based instruction considering the fundamental role these skills play in acquiring a broad array of academic skills beyond prior mathematics performance. Gains in expressive and receptive vocabulary, on the other hand, might be capturing gains in development that happen across a wider variety of contexts, many of which may have little to do with how students are accessing instructional content. For instructional coaches, interventionists (i.e., mathematics or reading specialists, school psychologists), and teachers, measures like EBRS may represent more effective barometers of how students are able to access the core components of kindergarten instruction, suggesting that the gains students exhibit may show an increasing ability to access rigorous grade-level reading abilities, which then may transfer to mathematics regardless of their level of mathematics performance at the beginning of kindergarten. Nevertheless, the diminishing effect of these gains in mathematics development also suggest that, controlling for students' performance at kindergarten-entry, there are a significant amount of other individual differences that impact and accrue over time that mitigate the advantage of steeper kindergarten year EBRS gains. This may be a product of the shifting content and difficulty of the content areas covered in the ECLS-K: 2011 mathematics measure. Those early EBRS gains may help sustain students through first grade; however, gains in those basic skills, regardless if entry mathematics performance, may be not be enough on their own to help students continue to access rigorous mathematics content.

This interpretation would then imply that access to the language of mathematics hinges on not only students' prior gains but also the manner by which those prior gains are sustained and reinforced. Combined with the effect of working memory ability, a rich, language-heavy mathematics environment may help sustain early basic reading gains to buffer students through the increasingly linguistically complex environment of mathematics. Because of the sequential nature of mathematics content domains (e.g., the emphasis on fraction/ratio and algebra in intermediate elementary) and thus the changing linguistic context of mathematics, such early gains in basic English reading skills likely need to be sustained as the target of mathematical language shifts. For example, if practitioners examined a student's yearly growth in kindergarten and saw they gained quickly enough to reach grade level content, that does not imply that they will be able to sustain that level of access to instruction consistently in subsequent years. The advantages from early basic reading gains that may accrue into increased instructional access does not imply that students will naturally be able to sustain grade-level content access. As such, practitioners can scaffold prior gains by engaging in practices such as dynamic assessment among ELLs (e.g., Orosco et al.,, 2011; Orosco, 2014), which can help assess students' content-area and grade-level knowledge when provided when the appropriate instructional supports, thus ruling out the potential for individual differences to impact the extent to which they can access grade-level content. Importantly, these implications may be helpful for students regardless of levels of other factors that are critical to instructional access (e.g., attentional focus, working memory, cognitive flexibility, prior mathematics ability). Indeed, because the effect of early basic reading gains can be interpreted as the effect holding constant prior mathematics performance, early language gains in the form of the knowledge about the structure and components of the English language (as opposed to vocabulary knowledge), may serve as a key lever for change, provided that interventionists and teachers help explicitly support ELLs through the complex linguistic structure of mathematics with evidence-based practices, such as the problem verbalization components of explicit instruction. Doabler, Clarke, Kosty, et al. (2016) established the effectiveness of these components for ELLs students at tier 1 in kindergarten, while Doabler et al. (2019) similarly found evidence for explicit instruction among ELLs at tier 2. Though much more empirical and qualitative work must be established, these techniques show promise in helping ELLs navigate the language of mathematics. With potential evidence that gains in basic English reading skills may at least temporarily benefit ELLs, further attention should be given to scaffolding the language of mathematics to further strengthen the connections between mathematical language, basic reading skills, and English language proficiency.

Reading (Chen & Chaloub-Deville, 2016) and language (Chow & Ekholm, 2019; LeFevre et al., 2010; Powell et al., 2017; Vukovic & Lesaux, 2013) have evidenced strong connections to later mathematics achievement, and while not directly hypothesized a priori with the measure used (EBRS), the results presented herein present a similar pattern. Relationships to language aside, the within-domain development of mathematics requires further attention in practice to untangle the multiple exo and endogenous factors contribute to mathematics abilities among ELLs. Specifically, the nature of the measures used in the current study suggest that while ELLs and EPSs exhibit similar patterns of inter and intraindividual change, this change is not occurring within similar subdomains of mathematics skills. The one-year lag in skill level at the end of kindergarten and subsequent skill growth ELLs demonstrate indicate that, while there are many unmeasured exogenous instructional and ecological factors impacting ELLs and EPSs between kindergarten and fourth grade, instruction as it occurs on average is not moving the needle on the gaps between ELLs and EPSs. Thus, though the quantity of inter and intraindividual change is similar between groups, this does not imply that similar developmental processes are occurring. As an example, between third and fourth grade, ELLs and EPSs would be predicted to change nearly the same amount based on their prior level, and their estimated change scores (across all three models considered) were highly similar. However, at this point, ELLs still performed around 0.80 *SD*s below EPSs. With the increasing focus on algebra by the third-grade round of assessment (Najarian et al., 2018b), coupled with the fact that the difference between ELLs and EPSs amounts to essentially one year of development, this subgroup of ELLs ends fourth grade with, on average, one year behind in their algebra, measurement, data analysis/probability, and numeracy abilities. With fifth grade being a potentially key year developmentally in preparing students for middle school, a much more effective dosage of instruction must be administered to prepare ELLs for an algebra, word-problem, and fraction-laden context of mathematics. In light of the other key developmental, psychosocial, emotional, and behavioral factors that make middle school a potentially challenging educational period, it is imperative to appropriately scaffold mathematics skills in order to buffer students through the changing stage of mathematics knowledge and educational context. This is potentially one explanation for the decelerating growth as ELLs gain more basic reading abilities: language may initially benefit students, but as the complexity of mathematics increases, prior mathematics skills may become increasingly important beyond the rate at which basic reading skills were acquired. Although few empirical examples of intermediate elementary mathematics interventions are currently published (though see Orosco, 2011 and Orosco et al., 2014), it is likely not a mystery as to how to best apply principles of explicit instruction and appropriate language and mathematics scaffolding to ELLs. Increasing progress-monitoring (formally through CBMs or more qualitatively in formative assessment) may help increase the ability to trace skill development and target specific skill deficits. Indeed, it would appear that as long as students acquire essential early numeracy skills, few other factors contribute to a significant achievement gap, which corroborates Jordan et al.'s (2007) findings that membership to a medium-steep or high-flat trajectory was most predictive of higher calculation outcomes. Progress-monitoring, opportunities to engage with grade-level content using the mathematical register (Kersaint et al., 2013) through explicit instruction (Jayanthi et al., 2008), and a combination of both native and second language skill reinforcement seem likely candidates to provide the necessary equity in access to instruction that Robinson-Cimpian et al. (2016) discuss. Yet given the consistent gaps over time, the most malleable, preventative lever would appear to be early numeracy skills based on the sheer amount of variance for which kindergarten-entry mathematics explains in the ELL-EPS achievement gap.

Many studies highlight the empirical and practical effectiveness of early numeracy, and the two studies focused on ELLs corroborate this evidence (Doabler, Clarke, Kosty, et al., 2016; Doabler et al., 2019), but far more insight from practitioners is needed to understand the real-world circumstances in which early numeracy instruction for lower ELP ELLs takes place. Practical perspectives in the intervention design process is critical to closing research-practice gaps as well as understanding the ecological and student-level mechanisms through which intervention effects are generated. Doabler, Clarke, Kosty, et al. (2016) recommend that an emphasis on vocabulary and the language of mathematics can likely bolster ELL success in tier-1 early numeracy instruction; however, the manner by which teachers support such practices needs to be further understood to effectively test the effects and mechanisms of such a recommendation. Theoretically and empirically there is a basis for improving the mathematical language use of ELLs, but the instructional process through which that occurs, and under what circumstances, needs to be more fully addressed. In essence, the most relevant practical implication of these results is that more information needs to be gleaned from practitioners about how to best address the complexities of providing rigorous, grade-level instruction to ELLs given the cultural, economic, instructional, and policy barriers to doing so.

Future Directions

Future directions for this work are focused primarily on next steps for identifying mechanisms of mathematics growth across language groups and the heterogeneities in these relationships. Although Study 1 identified similar predictive capability of working memory across language groups, this does not imply that the similar variables of or magnitudes of the same variable mediate the relationship between working memory and mathematics between ELLs and EPSs. This hypothesis rests primarily on the different linguistic contexts (Van Rinsveld et al., 2016) of mathematics cognition and instruction and that, as a result, working memory may associate to mathematics through different cognitive or instructional pathways. Among ELLs specifically, a there is a similar need to assess mediators of the relationship from EBRS gains to mathematics. Working memory did not moderate the relationship of EBRS or ELP gains to mathematics growth as predicted; however, other cognitive and executive functioning factors may account for whom EBRS gains are most predictive in light of L1 and L2 learning and which cognitive factors generate such relationships effects.

In the current work, I used more typical multiple-group and mixture modeling approaches. In light of the findings and limitations of these approaches, future work may be able to more precisely examine ELL mathematics development through two strategies. First, the primary work of future studies should be to more critically examine the role of theory in assessing moderators and mediators. Hypotheses of the current work drew on both practical strategies used to assess ELP in schools and developmental/cognitive theory to assess predictors of mathematics development through more general approaches to assessing moderation (multiple-group invariance testing and creating interaction terms). These practical approaches included utilizing gains in ELP/EBRS growth to predict mathematics development given that such gains are typically used in schools to draw inferences on instructional response and the rate of ELP acquisition. Indeed, the EBRS measure mirrors assessments more akin to CBM in reading than the *pre*LAS, and findings for that measure should be interpreted as such. Nevertheless, similar measures in schools would be utilized to potentially make instructional provision decisions. Additionally, academic year growth is one of the core norming benchmarks among CBMs. Drawing on cognitive load theory (Sweller, 1988; Sweller et al., 2011), I hypothesized that working memory may moderate the strength of the relationship between ELP/EBRS and mathematics growth as lower working memory may increase the cognitive load of mathematics such that greater gains in language proficiency may then become more predictive of later mathematics performance and growth.

However, there are many design considerations limiting this perspective to assessing moderation of language gains, and alternative theories should be formulated to examine when differential changes relate to mathematics growth. These theories can draw heavily on early mathematics cognition and language development—especially from the triple-code model (Dehaene, 1992), biological mechanisms of mathematics cognition (Spelke, 2017), and potential bilingual benefits to executive function—to more precisely examine how much, for whom, and why (mediators) language levels and gains may relate to mathematics growth. The methodological strategies for such studies should directly relate to the theory in mind. For example, though the mixture modeling technique used in the current study has strong precedent in developmental research, it is limited to examining heterogeneity in the outcome variable. One technique that would be ideal to directly test drivers of heterogeneity in SEMs are the SEM tree (Brandmaier, Oertzen, McArdle, Lindenberger., 2013) and SEM forest (Brandamaier, Prindle, McArdle, & Lindenberger, 2016) approaches. As Jaccobucci et al. (2017) highlighted, the SEM tree approach can be equally effective in uncovering heterogeneity in multivariate relationships as mixture models. The process by which this occurs, however, differs significantly. In SEM tree analysis, covariate sets drive enumeration of trees consisting of differential relationships through recursive partitioning, maximizing between-partition differences and within-partition homogeneity (Brandmaier et al., 2013). Maximizing within-class homogeneity has similarity to the assumptions of latent class growth analysis (LGCA), which assumes no variability of latent growth factors within classes, though classes derived in LGCA are not conditional on covariates as in SEM trees. Many different assumptions must be made about an analysis using SEM trees, but it is an important technique in subgroup analysis that highly aligns with the methodological and theoretical questions of ELL mathematics development with relevance to RTI. The more general classification and regression tree (CART) analysis has been recently applied in a similar area of language and mathematics development to analyze differential relationships in regression models of language skills predicting mathematics outcomes among preschool-aged children (Purpura, Day, Napoli, & Hart, 2017). To that end, SEM tree analysis is an important extension in this area with its applicability to heterogenous multivariate relationships that directly test developmental theory or conditional processes.

Another important future direction is to continue to further research not only how much language matters in mathematics but what type (Chow & Jacobs, 2016; Chow & Eckholm, 2019). The role of language among EPSs continues to remain somewhat elusive, again in part due to how it's measured. Among ELLs, unlocking the specific components of language that matter most for mathematics may help reduce the language load and may improve the efficiency of instruction. This may also help identify ways to help students across the ELP continuum access the curriculum. Studying ELP as a broad construct, comprised of many aspects of reading and language, in relationship to mathematics would likely benefit from more nuanced and granular measurement to specify how and to what extent L2 components matter in mathematics and how L1 facilitates or impedes this relationship. Chow and Jacobs (2016) highlighted some of the shortcomings of language measurement specifically in relationship to fractions. L2 vocabulary is undoubtedly important for accessing curriculum, yet in relation to mathematics, facilitating understanding of the structure of language and how that structure relates to mathematics learning may help clarify the processes through which ELLs access mathematics knowledge in L2.

Conclusions

Factors that contribute to positive mathematics development are highly multidimensional. Cultural, socioeconomic, and systemic (e.g., district practices for ELL exit status) factors that impact students early on as well as consistently across schooling are crucial to structuring screening systems, data-based decision making, and monitoring intervention progress. This work shows that development occurs similarly for ELLs and EPSs (potentially more so than is typically suspected); however, individual differences in mathematics abilities, language proficiency and prior development interact with the ecology of schools, homes, and mesosystems (e.g., school-home interactions) differently for different students. For example, cultural factors may guide the manner by which parents interact with schools and view the relationship of schools with their child's learning. Promoting sustainable mathematics skill development then requires practitioners and researchers to leverage students' strengths and the skills that maximize students' abilities to learn, which requires considering the students' ecology of development and, given the demographic factors of ELLs, the cultural factors that interact with school-based practices. Building capacity in multiple ecological systems can help strengthen the data corroboration process and triangulate information to increase sensitivity and specificity of risk identification and monitoring intervention response (Albers et al., 2013). The diverse needs of ELLs compared to EPS peers as well as within the ELL population necessitates systems-level coordination to promote monitoring of academic growth. This is especially important to consider in the light of the similarity of working memory in predicting mathematics development. There may be different processes for ELLs and EPSs in how this relationship cascades given the contextual relationships that may differentially impact ELL and EPS development. This is the basis for understanding the probabilistic epigenesis of development (Gottlieb, 2007) and resulting cascading effects of contexts on individual change patterns and changes across domains within individuals. As such, the similarities between groups should not be taken as similarities in within-group processes given the contexts of development prior to and throughout schooling.

Language-based mechanisms for mathematics interventions are crucial as well and serve as a core component of a number of evidence-based explicit-instruction interventions at tier 1 and 2. As Doabler, Nelson, and Clarke (2016) and Doabler, Clarke, Kosty, et al. (2016) emphasized, opportunities to meaningfully verbalize problem solving is critical in mathematics learning. As such, a combination of basic print knowledge and knowledge of the structure of L2 in combination with academic and social language capability likely comprises a fuller profile of the language components that relate to mathematics. Nevertheless, underlying these ELP areas is more domain-general language ability that likely facilitates mapping of exact to approximate quantities (Dehaene, 1992) and a route to mathematics competence (LeFevre et al., 2010). For ELLs, this domain-general language ability would likely facilitate both L1 and L2 proficiency. The lack of support for the hypothesis that ELP gains would predict mathematics development patterns indicates that much more research is required in both measurement and development to understand the role ELP gains play in subsequent mathematics growth, if any. The hypothesis that the rate of change of ELP facilitates mathematics development should not be ruled out as a potential developmental cascade with mathematics growth. Acquiring L2 typically relies on L1 proficiency, cognitive characteristics, and ecological processes (Halle et al., 2012), which would likely share significant variability with many other facilitators and protective factors of mathematics development.

To that end, the role of native language is a key factor in this cascade as well. Because L1 and L2 likely share many general language components, facilitating L1 is important in bolstering reading and mathematics abilities. Guglielmi (2012) supported this hypothesis among secondary-level ELLs between eighth and twelfth grade. Cárdenas-Hagan et al. (2007) also provided evidence for this cross-language transfer among young, Spanish-speaking ELLs. Beyond these developmental factors, and potentially equally important, is the cultural and social role L1 plays in student development, which likely also facilitate academic development. As Umansky (2016) described, the relationship between L1 and L2 in an instructional context may have significant social psychological implications for learning, for example, in dual-language classrooms where the balance between majority and minority languages may not be effectively managed. Hence, the cultural context and relevance of learning should remain a vital element to research the diverse needs of ELL mathematics learning, even when focusing on the cognitive factors that impact learning, as individual-by-context interactions promote and impede student success, health, and well-being.

In light of these important factors, the current work does not provide conclusive evidence regarding considerations for cognitive development, RTI implementation, or cultural/socioecological components of education. However, three primary conclusions can be drawn from this work. First, early language-group mathematics performance gaps do not necessarily differentially compound over time. The expected change individuals undergo between each year is not markedly different, especially when controlling for between-group differences, though the variance in such intraindividual change differs between groups (likely due simply to the significant differences in sample sizes). Second, variance in individual trajectories of English-based mathematics growth is quite noisy as opposed to differentially systematic. From an RTI perspective, this may suggest that projected growth trajectories may not be as effective indicators of academic risk since there is such significant variability in growth across levels of performance. The translation to practice of this is not direct, however. Rather, this indicates that the measurements used to assess student progress and instructional response (e.g., CBMs) need to undergo more rigorous evaluation to determine how to most effectively use them in practice to guide identification, intervention provision, and instruction. Significant research has been devoted to this in the summative and accountability assessment literature (e.g., Abedi, 2002; Solano-Flores, 2016). The intersection between mathematics cognition development, second language learning, CBM, and instruction/intervention has yet to undergo the same empirical examination.

Last, and perhaps more important to conclude from this work, is how research in this area ought to function moving forward. Fuchs and Fuchs (2017) highlighted significant limitations of the national RTI evaluation (Balu et al., 2015), which used a regression discontinuity design (RDD) design to evaluate RTI for reading in 146 schools. It found null effects for students' reading performance based on their potential tier 2 or 3 assignment with some outcome measures showing even negative effects. Yet, as Fuchs and Fuchs (2017) noted, the national evaluation had no information on whether students actually received intervention; causal effects were generated for students who scored differently, not who received differential dosages of instruction. Additionally, as Fuchs and Fuchs (2017) also noted, Balu et al. (2015) likely overestimated the cut score for tiered intervention eligibility in the RDD analysis. Methodological limitations notwithstanding, Fuchs and Fuchs advocate for a more critical eye as to why RTI does not generate the expected effects. Their point rests on the implementation of the policy and the provision of services within each tier. Weak evidence base for many interventions, few resources to effectively implement intervention to high fidelity, and assessment practices to efficiently and effectively identify at-risk students and monitor their progress all impact how inferences about RTI are drawn. Fuchs and Fuchs caution against inferences drawn from the discontinuity evaluation, as they may skew the progress towards understanding the science and evidence behind academic interventions for general and tiered instruction and how to most effectively implement the science in real-world settings.

These concerns and consideration for the evidence of RTI are especially poignant regarding mathematics assessment and intervention given the traditional dominance reading has had in the school psychology and education science literature (Balu et al.'s [2015] evaluation was conducted only on RTI for reading). Compounding this, then, is how to identify the science of mathematics intervention for ELLs and align this science with studies of mathematics cognition, development, bioecology, and education policy. RTI is inherently translational to practice; however, it can be strengthened through translation of content and theoretical domains within research in mathematics learning and cognition research. Establishing these connections and the malleable components to promote equitable provision of instruction and sustainable growth across language groups will ideally mitigate many of the factors that ostensibly diminish the perceived effectiveness of RTI. In turn, this interdisciplinary approach may help bolster the evidence for universal components of general classroom instruction, the concurrent and predictive validity of screening instruments (Albers & Martinez, 2015; Glover & Albers, 2007; Hall et al., 2019), the sensitivity of progress-monitoring tools to student growth (Ardoin &

Christ, 2008), and, in turn, the effectiveness and efficiency of both universal and tiered interventions.

References

Abedi, J. (2002). Standardized achievement tests and English language learners: Psychometric issues. *Educational Assessment, 8*, 231–257. doi: 10.1207/S15326977EA0803_02

Abedi, J., & Herman, J. (2010). Assessing English language learners' opportunity to learn mathematics: Issues and limitations. *Teachers College Record*, *112*, 723–746.

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*, 219–234. doi:10.1207/ S15324818AME1403_2

Ainley, M. & Ainley, J. (2015). Early science learning experiences: Triggered and maintained interest. In K. A. Renninger, M. Nieswandt, & S. Hidi (Eds.), *Interest in Math and Science Learning* (pp. 17–31). Washington, D.C.: American Educational Research Association.

Albers, C. A., & Martinez, R. S. (2015). *Promoting academic success with English language learners*. New York, NY: Guilford Press.

Albers, C. A., Mission, P. L., & Bice-Urbach, B. (2013). Considering diverse learner characteristics in problem-solving assessment. In R. Brown-Chidsey & K. J. Andren (Eds.), *A problem-solving approach* (2nd ed., pp. 101–122) New York, NY: Guilford Press.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5*. Washington, D.C.: American Psychiatric Association.

Andersson, U., & Lyxell, B. (2007). Working memory deficit in children with mathematical difficulties: A general or specific deficit? *Journal of Experimental Child Psychology, 96,* 197–228. doi: 10.1016/j.jecp.2006.10.001

Ardoin, S. P., & Christ, T. J. (2008). Evaluating curriculum-based measurement slope estimates using data from triannual universal screenings. *School Psychology Review*, *37*, 109–125.

Asparouhouv, T., & Muthen, B. (2014). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. *Mplus Web Notes: No. 15*. Retrieved from https://www.statmodel.com/examples/webnotes/webnote15.pdf.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, *49*(1), 5-28. doi: 10.1080/713755608

Baddeley, A., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47-89. doi: 10.1016/S0079-7421(08)60452-1

Bailey, D. H., Watts, T. W., Littlefield, A. K., & Geary, D. C. (2014). State and trait effects on individual children's differences in mathematical development. *Psychological Science*, *25*, 2017–2026. doi: 10.1177/0956797614547539

Baker, S. K., Lesaux, N., Jayanthi, M., Dimino, J., Proctor, C. P., Morris, J., . . . Newman-Gonchar, R. (2014). *Teaching academic content and literacy to English learners in elementary and middle school* (Report No. NCEE 2014-4012). Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Baker, S. K., Nelson, N. J., Stoolmiller, M., Paine, P. K., Turtura, J., Crone, D., & Fien, H. (2019). Intervening with struggling readers in seventh grade: Impact evidence from six school districts. *Journal of Research on Educational Effectiveness*, *11*, 479–506. doi: 10.1080/19345747.2018.1472357

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of response to intervention practices for elementary school reading* (NCEE: 2016-4000) Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Bardack, S. (2010). *Common ELL terms and definitions*. Washington, D.C.: American Institutes of Research. Retrieved from: https://www.air.org/sites/default/files/downloads/report/NEW_-xhttps://www.air.org/sites/default/files/downloads/report/NEW_-_Common_ELL_TERMS_AND_DEFINITIONS_6_22_10_0.pdf

Baroody, A. J., Feil, Y., & Johnson, A. (2007). An alternative reconceptualization of procedural and conceptual knowledge. *Journal for Research in Mathematics Education*, *38*, 115–131.

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology*, *65*, 229–235. doi: 10.1037/a0025406

Bialystok, E., & Martin, M. E. (2004). Attention and inhibition in bilingual children: Evidence from the Dimensional Change Card Sort task. *Developmental Science*, *7*, 325–339. doi: 10.1111/j.1467-7687.2004.00351.x

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*(4), 1186-1206. doi: 10.1111/cdev.12660

Brandmaier, A. M., Prindle, J. J., McArdle, J. J., & Lindenberger, U. (2016). Theory-guided exploration with structural equation model forests. *Psychological Methods, 21*, 566–582. doi:10.1037/met0000090

Brandmaier, A. M., von Oertzen, T., McArdle, J. J., & Lindenberger, U. (2013). Structural equation model trees. *Psychological Methods*, *18*, 71–86. doi: 10.1037/a0030001

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 793–828). Hoboken, NJ: John Wiley & Sons, Inc.

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*, 205–228. doi: 10.1080/87565640801982312

Burns, M. K. (2011). School psychology research: Combining theory and prevention science. *School Psychology Review, 40*, 132–139.

Byrnes, J. (1992). The conceptual basis of procedural learning. *Cognitive Development*, *7*, 235–257. doi: 10.1016/0885-2014(92)90013-H

Byrnes, J. P., & Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Developmental Psychology*, *27*, 777–786. doi: 10.1037/0012-1649.27.5.777

California Education for a Global Economy Initiative, Cal. S.B. 1174 (2013-14), Chapter 753 (Cal. Stat. 2014).

Cárdenas-Hagan, E., Carlson, C. D., & Pollard-Durodola, S. D. (2007). Cross-linguistic transfer of early literacy skills: The role of initial L1 and L2 skills and language of instruction. *Language, Speech, and Hearing Services in Schools, 38*, 249–259. doi: 10.1044/0161-1461(2007/026)

Chard, D. J., Clarke, B., Baker, S., Otterstedt, J., Braun, D., & Katz, R. (2005). Using measures of number sense to screen for difficulties in mathematics: Preliminary findings. *Assessment for Effective Intervention, 30*, 3–14. http://dx.doi.org/10.1177/073724770503000202

Chen, F., Chaloub-Deville, M. (2016). Differential and long-term language impact on math. *Language Testing, 33*, 577–605. doi: 10.1177/0265532215594641

Chen, X., & Li, Y. (2008). Language proficiency and mathematics learning. *School Science and Mathematics, 108*(3)*,* 90–93. doi: 10.1111/j.1949-8594.2008.tb17811.x

Chow, S. M., Grimm, K. J., Filteau, G., Dolan, C. V., & McArdle, J. J. (2013). Regime-switching bivariate dual change score model. *Multivariate Behavioral Research, 48*, 463–502. doi: 10.1080/00273171.2013.787870

Chow, J. C., & Jacobs, M. (2016). The role of language in fraction performance: A synthesis of literature. *Learning and Individual Differences*, *47*, 252–257. doi: 10.1016/j.lindif.2015.12.017

Chow, J. C., & Ekholm, E. (2019). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly*, *46*, 179–186. doi: 10.1016/j.ecresq.2018.02.01

Clarke, B., Baker., S., Smolkowski, S., & Chard, D.J. (2008). An analysis of early numeracy curriculum-based measurement: Examining the role of growth in student outcomes. *Remedial and Special Education*, *29*, 46–57. Doi: 10.1177/0741932507309694

Clarke, B., Doabler, C. T., Smolkowski, K., Baker, S. K., Fien, H., & Cary, M. (2016). Examining the efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Learning Disabilities, 49*, 152–165. doi: 10.1177/0022219414538514

Clarke, B., Doabler, C., Smolkowski, K., Kurtz Nelson, E., Fien, H., Baker, S. K., & Kosty, D. (2016). Testing the immediate and long-term efficacy of a tier 2 kindergarten mathematics intervention. *Journal of Research on Educational Effectiveness, 9*, 607–634. doi: 10.1080/19345747.2015.1116034

Clarke, B., Haymond, K., & Gersten, R. (2014). Mathematics screening measures for the primary grades. In R. J. Kettler, T. A. Glover, C. A. Albers, & K. A. Feeney-Kettler (Eds.). *Universal screening in educational settings: Evidence-based decision making for schools* (pp. 199-221). Washington, D.C.: American Psychological Association.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review*, *33*, 234–248.

Clarke, B., Smolkowski, K., Baker, S. K., Fien, H., Doabler, C. T., & Chard, D. J. (2011). The impact of a comprehensive tier 1 kindergarten curriculum on the achievement of students at risk in mathematics. *The Elementary School Journal*, *111*, 561–584. doi: 10.1086/659033

Codding, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of math basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice*, *26*, 36–47. doi: 10.1111/j.1540-5826.2010.00323.x

Compton, D., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, *98*, 394–409. doi: 10.1037/0022-0663.98.2.394

Cragg, L., & Gillmore, C. (2014). Skills underlying mathematics: The role of executive functioning in the development of mathematics proficiency. *Trends in Neuroscience and Education*, *3*, 63–68. doi: 10.1016/j.tine.2013.12.001

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition, 162,* 12–26.

doi: 10.1016/j.cognition.2017.01.014

Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review*, *34*, 344–377. doi: 10.1016/j.dr.2014.10.001

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49,* 222–251.

Dee, T., & Penner, E. (2017). The causal effects of cultural relevance: Evidence from an ethnic studies curriculum. *American Educational Research Journal, 54*, 127–166. doi: 10.3102/0002831216677002

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics.* New York, NY: Oxford University Press.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences, 21*, 355–361. doi: 10.1016/S0166-2236(98)01263-6

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology, 20*, 487–506. doi: 10.1080/02643290244000239.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science*, *284*(5416), 970–974. doi: 10.1126/science.284.5416.970

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science*, *38*, 105–134. doi: 10.1007/s11251-009-9110-0

Doabler, C. T., Clarke, B., Kosty, D. B., Baker, S. K., Smolkowski, K., & Fien, H. (2016). Effects of a core kindergarten mathematics curriculum on the mathematics achievement of Spanish-speaking English language learners. *School Psychology Review, 45*, 343–361. doi: 10.17105/SPR45-3.343-361

Doabler, C. T., Clarke, B., Kosty, D., Smolkowski, K., Nelson, E., Fien, H., & Baker, S. K. (2019). Building number sense among English learners: A multisite randomized controlled trial of a tier 2 kindergarten mathematics intervention. *Early Childhood Research Quarterly*, *47*, 432–444. doi: 10.1016/j.ecresq.2018.08.004

Doabler, C. T., Nelson, E., & Clarke, B. (2016). Adapting evidence-based practices to meet the needs of English learners with mathematics difficulties. *Teaching Exceptional Children*, *48*, 301–310. doi: 10.1177/0040059916650638

Driver, M. K., & Powell, S. R. (2017). Culturally and linguistically responsiveness schema instruction: Improving word problem solving for English language learners with mathematics difficulty. *Learning Disability Quarterly*, *40*, 41–53. doi: 10.1177/073/1948716646730

Duncan, S. E., & De Avila, E. A. (1998). *PreLAS 2000*. Monterey, CA: CTB/McGraw-Hill.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446. doi: 10.1037/0012-1649.43.6.1428.

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, *50*, 2417–2425. doi: 10.1027/a0037996

Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: Guilford.

Enders, C. K., & Gottschall, A. C. (2011). Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling, 18*(1), 35-54. doi:

10.1080/10705511.2011.532695

Enders, C. K., & Tofighi, D. (2008). The impact of misspecifying class-specific residual variances in growth mixture models. *Structural Equation Modeling, 15*, 75–95. doi: 10.1080/10705510701758281

Faulkner-Bond, M., Waring, S., Forte, E., Crenshaw, R. L., Tindle, K., & Belknap, B. (2012). *Language instruction education programs (LIEPs): A review of the literature*. Washington, D.C.: U.S. Department of Education.

Ford, D.H., & Lerner, R.M. (1992). *Developmental systems theory*. Newbury Park, CA: Sage Publications.

Fuchs, D., & Fuchs, L. S. (2017). A critique of the national evaluation of response to intervention: A case for simpler frameworks. *Exceptional Children*, *83*, 255–268. doi: 10.1177/0014402917693580

Fuchs, L. S., Compton, D. L., Fuchs, D., Paulsen, K., Bryant, J. D., & Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*, 493–513. doi: 10.1037/0022-0663.97.3.493

Fuchs, L. S., & Fuchs, D. (2007). A model for implementing responsiveness to intervention. *Teaching Exceptional Children*, *39*, 14–20.

Fuchs, L. S., Fuchs, D., & Gilbert, S. K. (2019). Does severity of students' pre-intervention math deficits affect responsiveness to generally effect first-grade intervention? *Exceptional Children*, *85*, 147–162. doi: 0.1177/0014402918782628

Fuchs, L. S., Powell, S. R., Cirino, P. T., Schumacher, R. F., Marrin, S., Hamlmet, C. L…Changas, P. C. (2014). Does calculation or word-problem instruction provide a stronger route to prealgebraic knowledge? *Journal of Educational Psychology, 106*, 990–1006. doi: 10.1037/a0036793

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, M., Malone, A….Changas, P. (2014). Does working memory moderate the effect of fraction intervention? An aptitude-treatment interaction. *Journal of Educational Psychology*, *106*, 199–514. doi: 10.1037/a0034341

Garrett, R. S. (2010). *Multilingualism, mathematics achievement, and language instructional policy* (Doctoral dissertation). Retrieved fromProQuest Dissertations Publishing. (ATT 3419635)

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*, 177–190. doi: 10.1037/0012-1649.40.2.177

Geary, D. C., Hoard, G. M., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104*, 206–223. doi: 10.1037/a0025398

Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: An eight-year longitudinal study. *Journal of Educational Psychology*, *109*, 680–693. doi: 10.1037/edu0000159.

Gelman, R., & Butterworth, B. (2005). Number and language: How are they related? *TRENDS in Cognitive Science, 9*(1)*,* 6–10. doi: 10.1016/j.tics.2004.11.004

Gersten, R., Baker, S. K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R.

(2007). *Effective literacy and English language instruction for English learners in the elementary grades: A practice guide* (NCEE 2007-4011).. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45,* 117–135. doi: 10.1016/j.jsp.2006.05.005

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association*, *56*(296), 841–868.

Gottlieb, G. (2007). Probabilistic epigenesis. *Developmental Science*, *10*, 1–11. doi: 10.1111/j.1467-7687.2007.00556.x

Gottschal, A. C., West, S. G., & Enders, C. K. (2012). A comparison of item-level and scale-level multiple imputation for questionnaire batteries. *Multivariate Behavioral Research*, *47*, 1–25. doi: 10.1080/00273171.2012.640589

Greenfader, C. R. M. (2017). *The role of executive function in the early academic achievement of Hispanics: A national study* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (ATT 10638158)

Griffiths, P. E., & Tabery, T. (2013). Chapter three – developmental systems theory: What does it explain and how does it explain it? *Advances in Child Development and Behavior*, *44*, 65-94. doi: 10.1016/B978-0-12-397947-6.00003-9

Grimm, K. J., & Ram, N., & Estabrook, R. (2010). Non-linear structured growth mixture models in M*plus* and OpenMX. *Multivariate Behavioral Research*, *45*, 887–909. doi: 10.1080/00273171.2010.531230

Grundy, J. G., & Timmer, K. (2017). Bilingualism and working memory capacity: A comprehensive meta-analysis. *Second Language Research*, *33*, 325–340. doi: 10.1177/0267658316678286

Guglielmi, R.S. (2012). Math and science achievement in English language learners: Multivariate latent growth modeling of predictors, mediators, and moderators. *Journal of Educational Psychology, 104*, 580–602. doi: 10.1037/a0027378

Gutiérrez, R. (2002). Beyond essentialism: The complexity of language in teaching mathematics to Latina/o students. *American Educational Research Journal, 39,* 1047–1088. doi: 10.3102/000283120390041047

Hall, G. J., Moore, E. C., McMackin, M., Markham, M. A., & Albers, C. A. (February 2019). *Predictive validity of elementary mathematics screeners for English language learners.* Paper presented at the 2019 National Association of School Psychologists Conference. Atlanta, GA.

Halle, T., Hair, E., Wandner, L., McNamara, M., & Chien, N. (2012). Predictors and outcomes of early veruss later English language prociciency among English language learners. *Early Childhood Research Quarterly*, *27*, 1–12. doi: 10.1016/j.ecresq.2011.07.004

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making* (NCEE 2009-4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Hartano, A., Yang, H., & Yang, S. (2018). Bilingualism positively predicts mathematical competence: Evidence from two large-scale studies. *Learning and Individual Differences*, *61*, 216-227. doi: 10.1016/j.lindif.2017.12.007

Hassinger-Das, B., Jordan, N. C., & Dyson, N. (2015). Reading stories to learn math:

Mathematics vocabulary instruction for children with early numeracy difficulties. *The Elementary School Journal*, *116*, 242–264. doi: 10.1086/683986

Hong, S., & You, S. (2012). Understanding Latino children's heterogenous academic growth trajectories: Latent growth mixture modeling approach. *The Journal of Educational Research*, *105*(4), 235-244. doi: 10.1080/00220671.2011.584921

Jaccobucci, R., Grimm, K., & McArdle, J. J. (2017). A comparison of methods for uncovering sample heterogeneity. Structural equation model trees and finite mixture models. *Structural Equation Modeling*, *24*, 270–282. doi: 10.1080/10705511.2016.1250637

Jayanthi, M., Gersten, R., Baker, S. (2008). *Mathematics instruction for students with learning disabilities or difficulty learning mathematics: A guide for teachers*. Portsmouth, NH: RMC Research Corporation, Center on Instruction. Retrieved from: https://files.eric.ed.gov/fulltext/ED521882.pdf

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice, 22*, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x

Jordan, N. C., Kaplan, D., Oláh, L. N., Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development, 77*, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x

Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology, 104*(3), 647–660. doi: 10.1037/a0029018

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance for education policy. *Peabody Journal of Education*, *77*, 189–215. doi:

10.1207/S15327930PJE7704_9

Kaplan, D. (2008). An overview of Markov chain methods for the study of stage-sequential developmental processes. *Developmental Psychology*, *44*, 457–467. doi: 10.1037/00121649.44.2.457

Kaplan, D. (2009). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.

Kaplan, D. (2014). *Bayesian statistics for the social sciences.* New York, NY: Guilford Press.

Kaplan, D., & Walpole, S. (2005). A stage-sequential model of reading transitions: Evidence from the Early Childhood Longitudinal Study. *Journal of Educational Psychology*, *97*, 551–563. doi: 10.1037/0022-0663.97.4.551

Kaufmann, L., & von Aster, M. (2012). Diagnostik und intervention bei rechenstörung. *Deutsches Arzteblatt International*, *109*, 767–778. doi: 10.3238/arztebl.2012.0767

Kersaint, G., Thompson, D., & Petkova, M. (2013). *Teaching Mathematics to English language learners*. New York, NY: Routledge.

Kieffer, M. J., & Thompson, K. D. (2018). Hidden progress of multilingual students on NAEP. *Educational Researcher*, *47*, 391–398. doi: 10.3102/0013189X18777740

Kohli, N., Sullivan, A. L., Sadeh, S. S., & Zopluoglu, C. (2015). Longitudinal mathematics development of students with learning disabilities and students without disabilities: A comparison of linear, quadratic, and piecewise linear mixed effects models. *Journal of School Psychology*, *53*, 105–120. doi: 10.1016/j.jsp.2014.12.002

Korkmanz, S., Goksuluk, D., & Zararsiz, G. (2014). MVN: An R package for assessing multivariate normality. *The R Journal*, *6*, 151–162.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8-9-year-old students. *Cognition*, *93*, 99–125. doi: 10.1016/j.cognition.2003.11.004

Lee, K., & Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *Journal of Educational Psychology, 108*, 869–882. doi: 10.1037/edu0000090

LeFevre, J., Fast, L., Schwarchuk, Smith-Chant, B.L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Development*, *81*, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x

Lerner, R. M. (1978). Nature, nurture, and dynamic interactionism. *Human Development, 21*(1), 1–20. doi: 10.1159/000271572

Lerner, R. M., Leonard, K., Fay, K., & Issac, S. S. (2011). Continuity and discontinuity in development across the life span: A developmental systems perspective. In K. L. Fingerman, C. A. Berg, J. Smith, & T. C. Antonucci (Eds.), *Handbook of life-span development* (pp. 141–182). New York, NY: Springer.

Li, Y., & Geary, D. C. (2013). Developmental gains in visuospatial working memory predict gains in mathematics achievement. *PLoS ONE*, *8*(7), ARTIcle ID e70160.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: a meta-analysis. *Psychological Bulletin Review*, *21*, 861–883. doi: 10.3758/s13423-013-0565-2

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities, 41,* 451–459. doi: 10.1177/0022219408321126

Loeb, S., Soland, J., & Fox, L. (2014). Is a good teacher a good teacher for all? Comparing value-added of teachers with their English learners and non-English learners. *Educational Evaluation and Policy Analysis*, *36*, 457–475. doi: 10.3102/0162373714527788

Lukasik, K., Lehtonen, M, Soveri, A., Waris, O., Jylkkä, J. & Laine, M. (2018). Bilingualism and working memory performance: Evidence from a large-scale online study. *PLoS ONE*, *13*(11), 1–16. doi: 10.1371/journal.pone.0205916

Lumley, T. (2004) Analysis of complex survey samples. *Journal of Statistical Software*, *9*(1), 1–19.

Lumley, T. (2018). *survey: Analysis of complex survey samples*. R package version 3.34.

Marsh, L. G., & Maki, R. H. (1976). Efficiency of operations in bilinguals as a function of language. *Memory and Cognition*, *4*, 459–464.

Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology*, *22*(3), 491-495. doi: 10.1017/S0954579410000222

Master, B., Loeb, S. Whitney, C., & Wyckoff J. (2012). Different skills? Identifying differentially effective teachers of English language learners. *The Elementary School Journal, 117,* 261–284. doi: 10.1086/688871

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9), 1-8. doi: 10.1371/journal.pone.0023749

McArlde, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*(1), 577-605. doi: 10.1146/annurev.psych.60.110707.163612

McArdle, J. J., & Nesselroade, J. R. (2014). *Longitudinal data analysis using structural equation models.* Washington, D.C.: American Psychological Association.

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer,. . . Bullock Mann, F. (2018). *The condition of education 2018* (NCES 2018-144). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018144.

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology, 74,* 240–260. doi:10.1006/jecp.1999.2516

Moll., K., Göbel, S. M., Gooch, D., Landerl, D., & Snowling, M. J. (2016). Cognitive risk factors for specific learning disorder: Processing speed, temporal processing, and working memory. *Journal of Learning Disabilities*, *49*, 272–281. doi: 10.1177/0022219414547221

Murphy, D. (2014). The academic achievement of English language learners: Data for the U.S. and each of the states. *Child Trends*. Retrieved from: https://www.childtrends.org/wp-content/uploads/2015/07/2014-62AcademicAchievementEnglish.pdf

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide*. (8th ed.). Los Angeles, CA: Author.

Najarian, M., Tourangeau, K., Nord, C., and Wallner-Allen, K. (2018a). Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), Kindergarten Psychometric Report (NCES 2018-183). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Najarian, M., Tourangeau, K., Nord, C., and Wallner-Allen, K. (2018b). Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), First- and Second-Grade Psychometric Report (NCES 2018-183). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education:

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. G., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Child Research Quarterly*, *36*, 550–560. doi: 10.1016/j.ecresq.2016.02.003

Noël, M. P., & Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: An explanation based on a double exact-and-approximate number representation model. *Frontiers in Human Neuroscience*, *5*, 1–4. doi: 10.3389/fnhum.2011.00165

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling*, *21*, 439–454.

Nys, J., Content, A., & Leybaert, J. (2013). Impact of language abilities on exact and approximate number skills development: Evidence from children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *56*, 956–970. doi: 10.1044/1092-4388

Oberski, D. (2014). lavaan.survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software*, *57*(1), 1-27. doi: 10.18637/jss.v057.i01

Orosco, M. J. (2014). Word problem strategy for Latino English language learners at risk for math disabilities. *Learning Disabilities Quarterly*, *37*, 45-53. doi: 10.1177/0022466911416248

Orosco, M. J., Swanson, H. L., O'Conner, R., & Lussier, C. (2011). The effects of dynamic strategic math on English language learners' word problem solving. *The Journal of Special Education*, *47*, 96-107. doi: 10.1177/0022466911416248

Petscher, Y., Koon, S., & Herrera, S. (2016). Using latent change score analysis to model co-developmental in fluency skills. In K. D. Cummings & Y. Petscher (Eds.), *The Fluency*

Construct (pp. 333–364). New York, NY: Springer.

Powell, S. R., Cirino, P.T., & Malone, A. S. (2017). Child-level predictors of responsiveness to evidence-based mathematics intervention. *Exceptional Children*, *83*, 359–377. doi: 10.1177/0014402917690728

Purpura, D. J, & Ganley, C. M. (2014). Working memory and language: Skill-specific or domain-general relations to mathematics? *Journal of Experimental Child Psychology*, *122*, 104–121. doi: 10.1016/j.jecp.2013.12.009

Purpura, D. J., & Logan, J. A. R. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, *51*, 1717–1724. doi: 10.1037/dev0000055

Purpura, D. J., Napoli, A. R., Wehrspann, E. A., Gold, Z. S. (2017). Causal connections between mathematical language and mathematical knowledge: A dialogic reading intervention. *Journal of Research on Educational Effectiveness*, *10*(1), 116-137. doi: 10.1080/19345747.2016.1204639

Purpura, D. J., & Reid, E. E. (2016). Mathematics and language: individual and group differences in mathematical language skills in young children. *Early Childhood Research Quarterly*, *36*,. 259-268. doi: 10.1016/j.ecresq.2015.12.020 0885-2006

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the children's behavior questionnaire. *Journal of Personality Assessment*, *87*, 103–113.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ram, N., & Grimm, K. (2009). Growth mixture modeling: A method for identifying differences in longitudinal change among unobserved groups. *International Journal of Behavior*

Disorders, *33*, 565–576. doi: 10.1177/0165025409343765

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

Resnick, I., Jordan, N. C., Hansen, N., Rajan, V., Rodrigues, J., Siegler, R. S., & Fuchs, L. S. (2016). Developmental growth trajectories in understanding of fraction magnitude from fourth through sixth grade. *Developmental Psychology*, *52*, 746–757. doi: 10.1037/dev0000102

Rinne, L. F., Ye, A., & Jordan, N. C. (2017). Development of fraction comparison strategies: A latent transition analysis. *Developmental Psychology*, *53*, 713–730. doi: 10.1037/dev0000275

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*, 175–189.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology, 93*, 346–362. doi: 10.1037//0022-0663.93.2.346

Roberts, G., & Bryant, D. (2011). Early mathematics achievement trajectories: English-language learner and native English-speaker estimates, using the Early Childhood Longitudinal Survey. *Developmental Psychology, 47*, 916–930. doi: 10.1037/a0023865

Robinson-Cimpian, J. P., Thompson, K. D., & Umansky, I. M. (2016). Research and policy considerations for English learner equity. *Policy Insights from the Behavioral and Brian Sciences, 3*, 129–137.

Rolstad, K., Mahoney, K., & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy, 19,* 572–594. doi: 10.1177/0895904805278067

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. doi: 10.18637/jss.v048.i02

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592. doi: 10.1093/biomet/63.3.581

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys.* New York, NY: Wiley.

Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). *Technical Abstract (Woodcock-Johnson III Assessment Service Bulletin No. 2)*. Itasca, IL: Riverside Publishing.

Siegler, R. S. (1991). In young children's counting, procedures precede principles. *Educational Psychology Review*, *3*, 127–135. doi: 10.1007/BF01417924

Spelke, E. S. (2017). Core knowledge, language, and number. *Language Learning and Development*, *13*, 147–170. doi: 10.1080/15475441.2016.126

Spelke, E. S., & Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition, 78,* 45–88.

Stegmann, G., & Grimm, K. (2018). A new perspective on the effects of covariates in mixture models. *Structural Equation Modeling*, *25*, 167–178. doi: 10.1080/10705511.2017.1318070

Swanson, H. L., Kong, J., & Petcu, S. (2018). Math difficulties and working memory growth in English language learner children: Does bilingual proficiency play a significant role? *Language, Speech, and Hearing Services in Schools*, *49*, 379–394. doi:

Swanson, H. L., & Zheng, X. (2014). Memory difficulties in children and adults with learning disabilities. In H. L. Swanson, K. R. Harris, & S. Graham (Eds.), *Handbook of Learning Disabilities* (2nd ed., pp. 214-238). New York, NY: Guildford Press.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257–285. doi: 10.1207/s15516709cog1202_4

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.

Szucs, D., Devine, A., Soltesz, F., Nobes, A., & Garbiel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex, 49*, 2674–2688. doi: 10.1016/j.cortex.2013.06.007

Tang, Y., Zhang, W., Chen, K., Feng, S., Ji, Y., Shen, J., Reiman, E. R., & Yijun, L. (2006). Arithmetic processing in the brain shaped by cultures. *Proceedings of the National Academic of Sciences*, *103*, 10775–10780. doi: 10.1073/pnas.0604416103

Tourangeau, K., Nord, C., Le, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L...Mulligan, G. M. (2015a). *Early childhood longitudinal study, kindergarten class of 2010-11 (ECLS-K:2011) user's manual for the ECLS-K:2011 kindergarten base year data file and electronic codebook, public version*. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Tourangeau, K., Nord, C., Le, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L...Mulligan, G. M. (2015b). *Early childhood longitudinal study, kindergarten class of 2010-11 (ECLS-K:2011) user's manual for the ECLS-K:2011 kindergarten-first grade data file and electronic codebook, public version*. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Tourangeau, K., Nord, C., Le, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L...Mulligan, G. M. (2018). *Early childhood longitudinal study, kindergarten class of 2010-11 (ECLS-*

K:2011) user's manual for the ECLS-K:2011 kindergarten-fourth grade data file and electronic codebook, public version. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

Umansky, I. M. (2016). To be or not to be EL: An examination of the impacts of classifying students as English learners. *Educational Evaluation and Policy Analysis*, *38*, 714–737. doi: 10.3102/0162373716664802

Umansky, I. M. (2018). According to plan? Examining the intended and unintended treatment effects of EL classification in early elementary and the transition to middle school. *Journal of Research on Educational Effectiveness*, *11*(4), 588-621. doi: 10.1080/19345747.2018.1490470

U.S. Department of Education, National Center for Education Statistics. (2018a). Table 204.20: English language learner (ELL) students enrolled in public elementary and secondary schools, by state: Selected years, Fall 2000 through Fall 2016. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of Educational Statistics* (2018 ed.). Retrieved from: https://nces.ed.gov/programs/digest/d18/tables/dt18_204.20.asp

U.S. Department of Education, National Center of Education Statistics. (2018b). Table 204.27: *E*nglish language learner (ELL) students enrolled in public elementary and secondary schools, by home language, grade, and selected student characteristics: Selected years, 2008-09 through fall 2016. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of Educational Statistics* (2018 ed.). Retrieved from: https://nces.ed.gov/programs/digest/d18/tables/dt18_204.27.asp

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). *mice*: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. Retrieved from: https://www.jstatsoft.org/v45/i03/

Van Rinsveld, A., Dricot, L., Guillaume, M., Rossion, B., & Schiltz, C. (2017). Mental arithmetic in the bilingual brain: Language matters. *Neuropsychologia*, *101*, 17–29. doi: 10.1016/j.neuropsychologia.2017.05.009

Van Rinsveld, A., Schiltz, C., Brunner, M., Landerl, K., & Ugen, S. (2016). Solving arithmetic problems in first and second grade: Does the language context matter? *Learning and Instruction, 42*, 72–82. doi: 10.1016/j.learninstruc.2016.01.003

Venkatraman, V., Siong, S. C., Chee, M. W. L., & Ansari, D. (2006). Effect of language switching on arithmetic: A bilingual FMRI study. *Journal of Cognitive Neuroscience, 18*, 64–74. doi: 10.1162/089892906775250030

Vilorio, D. (2014). STEM 101: Intro to tomorrow's jobs. *Occupational Outlook Quarterly, Spring,* 2–12. Retrieved from: https://www.bls.gov/careeroutlook/2014/Spring/art01.pdf

Vukovic, R. K., & Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children's mathematical development. *Journal of Experimental Child Psychology, 115,* 227–244. doi: 10.1016/j.jecp.2013.02.002

Wang, W. (2017). *Initial language status and achievement trajectories among Hispanic students: Mediation through executive function* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (ATT 10682592)

Watts, T. W., Clements, D. H., Sarama, J., Wolfe, C. B., Spitler, M. E., & Bailey, D. H. (2017). Does early mathematics intervention change the processes underlying children's learning? *Journal of Research on Educational Effectiveness*, *10*(1), 96–115. doi: 10.1080/19345747.2016.1204640

Watts, T. W., Duncan, G. J., Chen, M., Claessens, A., Davis-Kean, P., Duckworth, K., Engel, M…Susperreguy, M. I. (2015). The role of mediators in the development of longitudinal achievement associations. *Child Development*, *86*, 1892–1907. doi: 10.1111/cdev.12416

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E…Gershon, R. C. (2013). Cognitive assessment using the NIH Toolbox. *Neurology*, *80*, 54–64. doi: 0.1212/WNL.0b013e3182872ded

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Widaman, K. F., & Thompson, J. S. (2003). On specifying the null model for incremental fit indices in structural equation modeling. *Psychological Methods*, *8*(1), 16-37. doi: 10.1037/1082-989X.8.1.16

Wilkinson, S. (2017). *Mathematics development of Spanish-speaking English language learners* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (ATT 10603152)

Woodcock, R. W, McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of cognitive abilities*. Itasca, IL: Riverside Publishing.

World-Class Instructional Design and Assessment. (2014). *ACCESS for ELLs 2.0 Summative Assessment*. Madison, WI: Author.

Zelazo, P. D. (2006). The dimensional change card sort (DCCS): A method of assessing executive function in children. *Nature Protocols*, *1*, 297–301. doi: 10.1038/nprot.2006.46