2022

Modeling Associations of English Proficiency and Working Memory with Mathematics Growth

Garret Hall and Craig Albers

Modeling Associations of English Proficiency and Working Memory with Mathematics Growth

Garret J. Hall[a] and Craig A. Albers[b]

[a]Department of Educational Psychology and Learning Systems, Florida State University

[b]Department of Educational Psychology, University of Wisconsin-Madison

Author Note

Correspondence regarding this manuscript should be sent to Garret J. Hall, Assistant Professor, Department of Educational Psychology and Learning Systems, Florida State University, 1114 W. Call St, 3204H Stone Building, Tallahassee, FL 32306, email: gjhall@fsu.edu

We have no known conflicts of interest to disclose.

We thank David Kaplan for feedback on the statistical methodology in this paper. 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 authors and do not represent views of the U.S. Department of Education.

This research was presented at the 2020 National Association of School Psychologists Convention in Baltimore, MD. This work is based on Garret J. Hall's PhD dissertation.

Abstract

Using kindergarten to fourth-grade data from the Early Childhood Longitudinal Study (2010-2011 cohort), we investigated systematic variability in English language learners' (ELLs; *n* = 303) mathematics growth as well as relations of kindergarten language growth and working memory (WM) to ELLs' mathematics growth. Using growth mixture modeling, only one class of growth emerged from ELLs' English mathematics growth from first through fourth grades. WM related to ELLs' English mathematics growth from Grades 1 to 4, as did kindergarten growth in English early literacy. We also investigated kindergarten to Grade 4 mathematics growth between ELLs and English-proficient students (EPSs; *n* = 4,711) using latent change score models and whether WM differentially predicted growth patterns. ELLs and EPSs did not exhibit markedly different growth patterns, and WM similarly predicted these patterns. Implications for future research as well as practical implications and limitations are discussed.

Keywords: *English language learners, mathematics, working memory, latent change score models, growth mixture modeling*

Impact Statement

There are minimal differences between ELLs' and EPSs' mathematics change patterns from kindergarten to fourth grade. Among only ELLs, English mathematics growth from Grades 1–4 was best characterized by a single, average nonlinear trajectory. Working memory and English early literacy growth predicted ELLs' mathematics development. These findings will assist school psychologists in understanding how ELLs' mathematics development transpires over time and what predicts these growth patterns. This information is important to consider when identifying prevention-focused assessment and instructional practices within multi-tiered systems of support.

Modeling Associations of English Proficiency and Working Memory with Mathematics Growth

English language learners (ELLs) face various language demands in their mathematics learning (Doabler et al., 2016), in addition to systemic equity barriers in general education (Robinsin-Cimpian et al., 2016) and special education (Sullivan, 2011). More clearly understanding developmental patterns in mathematical knowledge among ELLs (as compared to their English-proficient peers) is important to target potential performance gaps more effectively over time, identify students in need of additional support, and balance the unique linguistic factors impacting ELLs' learning within U.S. schools. These efforts require modeling of developmental trends of mathematics development to identify sources of systematic between- and within-person variability among ELLs. In addition, predictors of these sources of variability are key to consider in building prevention systems that rely on clear understanding of developmental processes (Kellam et al., 1999).

However, relatively little work to date has focused on disentangling variability in ELLs' mathematics development trends and what may predict these trends. This area of work is pertinent to understanding broad developmental trends that help formulate normative (national and local) comparisons among linguistically diverse students. These interindividual comparisons can also assist with contextualizing ipsative (i.e., within-person) comparisons that aid with instructional decision making. Longitudinal models of mathematics development that capture singular sources of variability in growth trends may not capture between- and within-person variance that are relevant to the practical application of broad-scale mathematics growth trends to prevention systems. This is especially the case when considering the unique factors related to ELLs' mathematics development as the distribution of growth rates among ELLs may be very heterogenous, and within-person growth trends may or may not differ substantially from that of English-proficient peers. Understanding these unique inter- and intragroup differences can better inform the field's capacity to understand mechanisms underlying mathematics growth among ELLs. Multi-tiered systems of support (MTSS) that recognize these unique differences in ELLs' mathematics development may be better equipped to tailor service delivery to their unique needs (Albers & Martinez, 2015).

Relatedly, key predictors of mathematics growth among ELLs have potential to operate differently compared with English proficient students (EPSs). Language proficiency and working memory (WM) have generally been underinvestigated as they relate to ELLs' mathematics development trends. Recent studies have addressed this gap more directly (e.g., Swanson et al., 2018, 2021), although both language (e.g., Chow & Eckholm, 2019) and WM (e.g., Geary et al., 2017) have been popularly examined as predictors among the general population. The purpose of this study was to address the research gap in modeling inter- and intraindividual variability in ELLs' mathematics growth and assess whether two key predictors (i.e., English language proficiency [ELP] development and WM) were relevant to ELLs' mathematics development across elementary school.

Measuring Patterns of Change

Different sources of change in mathematics among ELLs relative to EPSs are important to consider as students develop different competencies in mathematics throughout elementary school (e.g., early numeracy, arithmetic, prealgebraic thinking) in addition to language development. Typical studies comparing ELLs' mathematics growth with their English proficient or English monolingual peers compare mean trends and individual differences in trends over time (e.g., Roberts & Bryant, 2011). Whereas models of average growth and individual differences in growth rate are well-equipped to capture these mean trends, they do not account for intraindividual differences in change over time. Traditional growth curve models capture only differences in growth rate, but not how individuals change relative to themselves (over and above average growth rate over time). Models that account for this (e.g., latent change score models [LCSMs]; McArdle & Nesselroade, 2014) can advance understanding of ELLs' mathematics growth by not only modeling average trends and variability in trends across individuals but also the extent to which change occurs as a function of prior performance. When considering the goals of monitoring students' growth within MTSS (particularly at the broad universal screening level), separating these types of change patterns (intra- and interindividual) is particularly informed and highly aligned to the goals of data-based decision making (i.e., making normative and ipsative comparisons). The lack of research on these types of change patterns among ELLs indicates a need to consider LCSMs as a valid approach to modeling growth in terms of applicability to MTSS. Using LCSMs to compare ELLs to EPSs in terms of intra- and interindividual variability in change can further elucidate whether these sources of variability differ between groups. These group comparisons of developmental processes may help inform data-based decision-making within MTSS, such as making local versus national comparisons of performance and growth among ELLs (Albers & Martinez, 2015).

In addition to examining differences in growth between ELLs and EPSs, minimal research has investigated systematic heterogeneity in mathematics growth among ELLs. By systematic heterogeneity in growth, we mean growth trends and trend variability that are qualitatively and quantitatively distinct within a population. This contrasts with variability that is random in nature across the entire population, such as a random distribution of growth rates across all ELLs, which characterizes random differences in growth rates relative to a single, average growth rate. Systematic heterogeneity and variability in growth may imply that there are distinct subpopulations of growth that are initially unobserved (i.e., not recorded as initial data) but that can be uncovered through a specific kind of growth model. Studying systematic variability can help inform how schools might direct early assessment and instructional supports by identifying growth patterns suggestive of more or less academic risk. Indeed, one goal of MTSS is to capture students who may be differential at risk and tailor services appropriately.

Growth mixture modeling (GMM), which is one technique for modeling unobserved heterogeneity in growth, has been highly informative for modeling variation in growth that may be indicative of systematic, unobserved differential trends (e.g., Hong & You, 2012; Jordan et al., 2006, 2007). GMM allows for the possibility that average growth rates (and variances) might systematically vary across unobserved subpopulations (i.e., mixtures; Kaplan, 2002). This method has been infrequently applied to examining linguistically diverse students' mathematics growth (however, see Hong & You, 2012). It is important to identify systematic patterns in data that may help inform school psychologists' problem-solving and data-based decision-making regarding ELLs' performance in order to ensure equity in assessment and intervention provision through MTSS (Albers & Martinez, 2015). This is particularly important since many ELLs may be assessed in English, which may not reflect their actual mathematics strengths and needs. GMM relaxes the assumption that a single average trend and single variance characterizes the entire population, which may then allow more granularity in understanding ELLs' mathematics growth. Prior work focusing on kindergarten and early elementary students (regardless of ELL status) has found that three distinct trends might capture early numeracy growth (Jordan et al., 2006, 2007), including (a) a class that starts low and grows less, (b) a class that begins average and grows relatively more, and (c) a class that starts high and grows less. Hong and You (2012) found multiple classes of Latino students' mathematics growth between fall of kindergarten and spring of Grade 1. They found four distinct classes of growth, which generally matched other studies using GMM for mathematics growth (e.g., a class that starts high and stays high, classes that start in the middle of the distribution but show more growth, and a class that starts lower and grows less). However, research to date has not focused specifically on ELLs' mathematics growth and whether these general subclass patterns apply across elementary school, particularly when ELLs' are assessed in English.

If unobserved subclasses of growth among ELLs do exist, one practical implication is whether students remain in those latent classes over time. The basic growth mixture model can be further expanded to model whether students are likely to remain or stay within their given class. For example, students who are initially in a class that starts low and grows less may have some probability of switching to a class that started higher and tended to grow relatively more. This would suggest that there is systematic heterogeneity in growth but also that there are intraindividual transitions between subclasses. Furthermore, some students might be systematically more likely to remain in their growth class than others; those that have a higher tendency to switch classes are referred to as "movers," whereas those that tend to stay in classes can be referred to as "stayers" (Kaplan, 2008). In all, systematic heterogeneity in growth rates in addition to systematic heterogeneity in moving between growth classes can be highly informative for MTSS by elucidating who changes and when. This could potentially aid in understanding universal screening and monitoring growth by better identifying patterns of variability and whether these patterns are suggestive of differential risk.

Predictors of Change

In addition to modeling variability in growth among ELLs, predictors of change are also important to consider, as these predictors may help inform the key targets for assessment and screening within MTSS. Two key predictors of mathematics growth have been underexamined among ELLs are ELP development and WM. Prior studies have typically examined ELP by comparing language proficient and non-proficient students (Halle et al., 2012) or groups of students with different home languages (Roberts & Bryant, 2011). However, few studies have studied how rate of development of ELP relates to later mathematics skills, particularly when ELLs are assessed in English. Language is well-known to have a direct bearing on mathematics achievement among linguistically diverse children (Vukovic & Lesaux, 2013) and among all students, regardless of language status (Chow & Eckholm, 2019). Recent work has supported a mutualism theory of language and mathematics skills (Peng et al., 2020), which suggest that language assists in the encoding, storage, retrieval, and application of mathematical knowledge. However, the rate at which students acquire ELP has been underinvestigated among ELLs specifically as it relates to mathematics development, despite the rate of acquisition being a key target for ELP attainment and ELL exit status (Suhr et al., 2021; U.S. Department of Education, 2016). Other recent work demonstrates that early bilingual proficiency confers a benefit to both mathematics computation and WM development (Swanson et al., 2018), implying a role of early ELP acquisition as a mechanism to access mathematics content. This is consistent with what Peng et al. (2020) findings on language-mathematics mutualism would suggest about the function of language in shaping the storage and application of mathematical knowledge.

Similarly, WM has been consistently identified as a key predictor of mathematics achievement (Miller-Cotto & Byrnes, 2020) and mathematics development (Geary et al., 2017). Generally, WM is the capacity to maintain, manipulate, and use information within short-term storage (Gathercole et al., 2004). This domain-general capacity possesses clear relations to the cognitive processes underlying many mathematical functions, such as holding and manipulating verbal and symbolic number representations. Numerous studies have documented the importance of WM for development of mathematical skills (Peng et al., 2016; Raghubar et al., 2010). WM possesses key relations to early numeracy skills (Friso-van den Bos et al., 2014). Peng et al. (2016) provided an overall meta-analytic effect size of .35 between WM and mathematics skills, and they also showed that WM had the strongest relations to whole number calculation and word-problem solving. WM, however, has been underresearched as it applies to ELLs' mathematics development. Research recent has demonstrated a positive relation between bilingual proficiency and WM as well as mathematics computation growth (Swanson et al., 2018), yet more work is needed to understand how WM may operate among ELLs in predicting mathematics growth, although other work suggests WM in either English or Spanish predicts English mathematics problem-solving growth (Swanson et al., 2021). ELLs' developing ELP may interact with WM in predicting mathematics development throughout early elementary school given that the additional language load of second language development may shape how information is held and applied in executing mathematical functions (e.g., calculation, solving word problems); thus, lower WM might increase the cognitive load (Sweller, 1988; Sweller et al., 2011) of ELP in mathematics learning (Cummins, 1979).

Much of the work on WM suggests that content knowledge becomes increasingly important over time (Geary et al., 2017), but also that WM and achievement may bidirectionally relate across early grades (Miller-Cotto & Byrnes, 2020). Further constraints could be placed on ELLs' WM given the differential contextual and linguistic factors relating to their mathematics learning (i.e., the interaction between language proficiency and classroom instruction), potentially resulting in WM predicting mathematics growth trends differently among ELLs as compared with EPSs. Among ELLs specifically, WM could also moderate the extent to which ELP acquisition predicts mathematics development by constraining the mechanisms by which ELP acquisition predicts mathematics, such as access to grade-level content in English and retention, storage, and use acquired knowledge and ongoing problem-solving (e.g., Sweller, 1988). These predictions are critical to examine when adjusting for prior mathematics knowledge, thereby further explaining the extent to which language and WM uniquely predict mathematics growth trends, irrespective of existing mathematics proficiency.

The Present Work

Building upon recent studies (e.g., Swanson et al., 2018, 2021), more work in this area is needed to understand how WM and ELP interact in predicting mathematics development. In the present work, we addressed this research gap by using LCSMs to capture intra- and interindividual differences in change over time between language groups to more thoroughly investigate the sources of change across language groups and model heterogeneity in growth patterns among ELLs. We addressed two primary research questions.

Research Question 1

We sought to examine (a) patterns of change in mathematics development from kindergarten to fourth grade; (b) whether patterns of change differed between ELLs and EPSs; and (c) the extent to which WM may differentially predict mathematics performance status and growth rate. This question is divided into two parts:

(1a) 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 constant and proportional change?

For Research Question 1a, we were interested in whether ELLs and EPSs grew differently in mathematics as a result of different types of change (i.e., different interindividual change or different intraindividual change). We expected ELLs to exhibit more growth through early elementary than EPSs and then decelerate more than EPSs through later elementary grades, given shifting emphases on language-driven aspects of mathematical problem-solving over time (e.g., more complex language to express relationships). We tested the sources of change between groups (i.e., constant [linear growth rate] and proportional change [change in proportion to prior level over and above constant growth]) by considering multiple plausible models and identifying the model with the best relative fit. Together, testing differences in constant and proportional change help address whether the developmental process of mathematics may be different for ELLs as compared with EPSs.

(1b) To what extent does WM differ across ELLs and EPSs in uniquely predicting mathematics development patterns?

This question was exploratory as we were seeking to identify the model that indicated the best overall fit when freeing versus constraining the relationship between WM and growth factors. A model that fits better (based on the Bayesian information criterion [BIC]) when the paths between WM and growth factors are freed may suggest differential contributions of WM to mathematics performance.

Research Question 2

In Research Question 2, we examined whether there were unobserved subgroups of ELLs that show qualitatively different growth patterns, and if subgroups are present, whether students switch these classes over time. We also investigated the extent to which kindergarten ELP growth and WM predicted later mathematics performance and growth. We divided Research Question 2 into three parts as follows:

(2a) Are there unobserved, systematic, differential growth trajectories (i.e., mixtures) in mathematics development when the significant majority of ELLs are assessed in English (Grades 1–4)?

Prior work (Jordan et al., 2006, 2007) has pointed towards the presence of unobserved growth trends that may include students who start low and gain less (or vice versa) in addition to a more "typical" mean trajectory. Among ELLs, evidence of similar trajectories would suggest there are potential subpopulations of growth that may be at differential risk for sustained academic difficulties, whereas it is possible that other subpopulations might show differentially stronger mathematics trajectories. Because these subclasses (mixtures) are inherently unobserved (latent), model comparisons are used to detect the presence of mixtures (Kaplan, 2002). If these mixtures exist, we planned to test whether students switch classes over time and if there are systematic patterns of students who are more likely to move classes or stay in their initial class (Chow et al., 2013; Kaplan, 2008). Switching classes may additionally suggest that being in a "riskier" growth class is not deterministic over time; there may be some probability that certain students switch into a higher performing class, or vice versa. Again, this would have meaningful implications for screening students at each grade and monitoring growth over time by helping to uncover subgroups that may be at differential risk (or have differential strengths).

(2b) Do gains in ELP across kindergarten predict greater mathematics growth?

Part (b) (as well as [c] below) builds on part (a) by examining whether ELP growth in kindergarten relates to mathematics growth from Grades 1–4. This can be assessed regardless of the presence of mixture classes (the effect can be estimated within mixtures). We primarily predicted that ELP gains would positively predict both later growth level and trend given the role of ELP in accessing English-based mathematics information. However, we also recognized that ELP may not be a positive predictor of later growth given some evidence that early reading gains may predict slowed mathematics growth (Chow et al., 2013). A similar circumstance may occur with ELP, whereby schools emphasize language and literacy instruction with less attention to the same explicit focus on mathematics development and its key predictors (e.g., using mathematical language in explicit instruction; Doabler et al., 2016). Because we could not predict the presence of mixtures, we did not express hypotheses about the relationship of ELP across mixture classes.

(2c) Is ELP a stronger predictor of mathematics growth patterns among students with lower WM capacity? If mixture classes are present, how does this interaction operate across classes?

We predicted that ELP gains would be a stronger predictor of later mathematics development level and trend among students with lower WM capacity. Because we could not specifically predict the presence of mixtures, we did not have hypotheses about how WM might operate across classes.

Method

We used data from the public-use version of the Early Childhood Longitudinal Study: 2010-2011 Kindergarten Cohort Public-Use data file (ECLS-K: 2011). The full study included data from 18,174 students. At the time of this study, the ECLS-K: 2011 data included outcomes through the end of Grade 4.

Participants

We used ECLS-K: 2011 data from the fall and spring of kindergarten and the spring of first through fourth grade. We included only students who attended public schools from kindergarten to fourth grade, which is consistent with inclusion criteria in prior work (e.g., Roberts & Bryant, 2011). We defined ELLs as students whose parents reported that English was *not* the primary language at home and who were not fully English language proficient as assessed in the fall of kindergarten (n = 303 students, see ELP section). We defined EPSs (n = 4,711) as those who did not meet these criteria for ELL.

Measures

Mathematics

We use the vertically-scaled item response theory (IRT) scale scores of the mathematics achievement measure in the ECLS-K: 2011 (Cronbach's α = .92–.94 across waves). Najarian et al. (2018) reported no differential item functioning between the Spanish and English versions of the assessment. The distribution of items in the kindergarten through Grade 2 rounds of data collection weighs heavily toward number properties and operations, and by Grade 3, these items comprise only 40% of the items, with measurement, geometry, data and probability, and algebra gaining greater weight in the item pool (Najarian et al., 2018). Further discussion of this measure, including further information on reliability and validity, can be found in Najarian et al. (2018).

ELP

Oral ELP was assessed on the preschool language acquisition screener (*pre*LAS; Duncan & De Avila, 1998), which assesses receptive and expressive English language skills across two 10-item subtests (i.e., *Simon Says*, *Art Show*). The total raw score is 20, but Simon Says is weighted twice as much, so the total score is 30 (Tourangeau et al., 2018). *pre*LAS scores were highly reliable in the Fall and Spring of kindergarten within the ECLS-K: 2011 sample (Fall Cronbach's α = .91, Spring Cronbach's α = .89; Tourangeau et al., 2015). The ECLS-K: 2011 used a score of < 16 out of 30 to define limited ELP for eligibility for Spanish assessment (Tourangeau et al., 2018), and prior studies have used various criteria to determine ELL and/or limited ELP status (e.g., Roberts and Bryant's, 2011). Our definition of ELL was based on two factors: parent-report of home language in kindergarten (i.e., not English or English/unable to choose primary language) and *pre*LAS score in the fall of kindergarten. Based on the distribution of *pre*LAS scores among primary language (English/cannot choose or not English), we defined ELL in this study as those students who scored 22 or below on the fall kindergarten *pre*LAS and whose parents reported the primary home language was not English in kindergarten. Please see the Supplemental Materials for further description of ELL classification criteria and *pre*LAS score distributions (Table S1, Figure S1, Figure S2). For Research Question 2, we subtracted fall kindergarten *pre*LAS scores from spring kindergarten *pre*LAS scores to create a change score that we used as a predictor of mathematics growth.

Spring Kindergarten WM

WM was assessed using the Numbers Reversed subtest from the Woodcock-Johnson Test of Cognitive Abilities, Version III (WJ-III NR; Woodcock & McGrew, 2001). In this task, the child listened to and was subsequently asked to repeat backwards varying series of numbers of increasing length. This assessment showed adequate reliability in the norming sample (.87; Schrank et al., 2001). We used age-based standard scales in our analysis, which were scaled to have a *M* of 100 and *SD* of 15.

English Basic Reading Skills

All students received part of the ECLS-K reading measure in English called the English basic reading skills (EBRS) measure, which was a 20-question measure that included key early literacy areas such as phonemic awareness and letter-sound correspondence (Najarian et al.,

2018) scored on a 0–20 scale (Cronbach's α = .87; Tourangeau et al., 2015). For Research Question 1, we used EBRS as only a covariate to remove ELL-EPSs differences in baseline English reading skills that may be related to mathematics development and WM. We used the EBRS measure rather than the full reading measure because some students (n = 9) in our analytic sample did not have an IRT reading score even though they had EBRS scores. For Research Question 2, we manually calculated the EBRS change score from fall to spring of kindergarten (i.e., subtracted fall from spring scores) to use as a predictor of mathematics growth and to compare whether growth in EBRS predicted mathematics development differently than *pre*LAS growth. Because EBRS and preLAS included two overlapping items, we did not include *pre*LAS and EBRS in the same models for Research Question 2.

Covariates

For Research Question 1, we included covariates (in addition to EBRS) in our analyses to adjust for differences across ELL and EPS groups and to isolate the unique relationships of WM and kindergarten ELP growth to mathematics growth. Demographic covariates included sex, race/ethnicity, standardized socioeconomic status (M = 0, SD = 1 in the full ECLS-K: 2011 sample), age at first kindergarten entry (months), fall kindergarten type (full/half day), pre-kindergarten care arrangement (parent or non-parent), and whether the student was a first-time kindergartener. We also included as covariates teacher ratings of inhibitory control and attentional focus from the Child Behavior Questionnaire (CBQ; Putnam & Rothbart, 2006), cognitive flexibility (Dimensional Change Card Sort; Zelazo, 2006), and fall of kindergarten IRT mathematics performance. We included the same demographic, teacher rating, and direct assessment (cognitive flexibility and prior mathematics performance) covariates to address Research Question 2b and 2c.

Procedure

Data for this study, the ECLS-K: 2011 kindergarten to fourth-grade public release file, were publicly available through the U.S. Department of Education's National Center for Educational Statistics website (NCES). We used the public-use version of the data, which did not constitute human subjects research. See Supplemental Materials for additional information on the data and sampling design.

Missing Data

We used multiple imputation with predictive mean matching (PMM; Rubin, 1987) using R's (R Core Team, 2020) *mice* package (van Buuren & Groothuis-Oudshoorn, 2011) to handle remaining missingness. We mean-imputed the group means for missing mathematics, reading, WM, and cognitive flexibility data due to the minimal missingness (PMM performed poorly). More information on our use of PMM is provided in the Supplemental Materials.

Analyses

Research Question 1. We conducted all structural equation modeling in Mplus Version 8.5 (Múthen & Múthen, 2012–2020). For Research Question 1, we used multiple-group latent change score modeling (MG-LCSM; McArdle & Nesselroade, 2014) to decompose group differences in mathematics development. LCSMs are composed of five primary parameters: mean and variance in level (intercept), linear growth rate and variance (constant change), and change relative to prior levels (proportional change)[1] simultaneously estimating proportional and constant change parameters yields a dual change model (McArdle & Nesselroade, 2014).

[1] In combination with proportional change, constant change could also be interpreted as the rate at which one reaches the peak (asymptote) of their developmental trajectory (Cáncer et al,, 2021). Because constant change varies across individuals, each individual would be predicted to reach their peak at different rates.

Different combinations of proportional and constant change values can produce decelerating trajectories (i.e., the predicted shape of development of mathematics), but the different combinations of values would suggest different processes of development. If the primary driver of change is proportional, it would potentially suggest that change is more so a function of prior knowledge than a constant function of time. Smaller proportional relative to constant change might suggest that the driver of change is more so a constant developmental process that varies across individuals. Using a multiple group model, we tested whether these inter- and intraindividual processes operate differently across EPSs and ELLs.

Research Question 2. The analytic technique for Research Question 2 consisted of multiple exploratory stages. We employed GMM to detect heterogeneity in average growth trends between the Spring of Grade 1 and the Spring of Grade 4 (when the majority of ELLs were assessed in English), totaling four measurement waves. We provide more information on our analytic approach for Research Question 2 in the Supplemental Materials. We used traditional latent growth curve modeling as the basis for GMM given that we were interested in modeling latent subclasses of average growth trends and growth trend variance. To address Research Questions 2b and 2c, we included Fall-Spring kindergarten *pre*LAS or EBRS gains, Spring kindergarten WM, and the interaction between WM and *pre*LAS/EBRS as predictors in the growth model, and we included additional covariates as described above in the covariates section.

Results

Table 1 presents descriptive statistics of continuous independent and dependent variables for the overall sample as well as each language group. Table 2 presents demographic characteristics of the sample and tests of baseline equivalence.

Research Question 1

(1a) 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 Constant and Proportional Change?

We first tested a series of unconditional MG-LCSMs to establish the initial model of change. We attended to the BIC as our primary measure of relative model fit. Smaller BIC values suggest a more favorable model (Kaplan, 2009). The BIC results for all models tested are presented in Table S3. The final model selected (Model 4) showed the most significantly reduced BIC relative to a single group model. Model 4, χ^2 [*df*] = 363.487 (16), evidenced mixed fit based on the root mean square error of approximation (RMSEA = .093), comparative fit index (CFI = .986), Tucker-Lewis index (TLI = .969), and standardized root mean squared residual (SRMR = .133).[2] Figure 1 displays the path diagram of the final model. ELLs and EPSs showed a nearly 0.80 *SD* gap in mathematics scores in kindergarten. In both groups, students who scored higher tended to grow more as indicated by the significant, positive proportional change values coupled with the positive constant change. However, this proportional effect tapered off by Grade 4. This suggests individuals' mathematics scores in Grade 3 became less predictive of change through Grade 4 (over and above constant change) for both ELLs and EPSs. Together, this produces a trajectory consistent with the commonly observed trend of decelerating mathematics growth throughout elementary school (e.g., Roberts & Bryant, 2011), although our use of MG-LCSMs allow us to examine this deceleration with more attention to intraindividual change (proportional change). In Figure S3, we show simulated trajectories based on the

[2] Our CFI and TLI calculations use a multiple-group intercept-only model with residuals constrained to equality over time and across groups as the null model (χ^2 [*df*] = 24,223.455 [35]) rather than the default null model, which is inappropriate for growth/latent change models (Widaman & Thompson, 2003).

parameters of this model; these trajectories better display the intra- and interindividual variability in growth captured in this model by displaying what observed data would look like had it been generated by the model depicted in Figure 1. We next considered a series of conditional models.

We included covariates into the MG-LCSM identified in the first step of model testing to analyze the extent to which the predictors of interest improved model fit and helped explain group differences. We first fit a model with all regression coefficients allowed to vary across groups; then we fit a model with all regression coefficients fixed to equality. We centered covariates at their survey-weighted grand means.

Figure 2 displays the MG-LCSM as a path model, and Table 3 displays the regression coefficients and fit indices. This path model, in which all regression coefficients were fixed to equality across groups, showed better fit based on the BIC (172144.972) than both the unconditional model and a conditional model with all regression effects freed across groups (BIC = 172314.811). Based on the change coefficients in Figure 2, EPSs showed consistent, significant proportional effects through Grade 4, although the direction changed after Grade 1. This suggests that higherperforming students at the end of kindergarten changed more through Grade 1, but performance level had the opposite relationship to change in the following waves. ELLs showed similar patterns after Grade 1, but prior performance was not significantly predictive of change beyond constant change in kindergarten or Grade 1. Controlling for covariates explained essentially all the ELL-EPS mathematics achievement gap at the end of kindergarten.

This model showed that the LCSM had a higher constant slope value and smaller proportional effect coefficients as compared with the unconditional model. This indicates growth occurred more so through a constant, linear process and less as a function of prior performance when the intercept and slope were conditioned on covariates. Figure 3 shows model-implied trajectories for the best-fitting unconditional model and the constrained-covariate conditional model. Few changes in the longitudinal trajectories between models emerged aside from the significantly reduced ELL-EPS performance gaps across time (see the Figure 3 note for important details on how we calculated the conditional trajectories). Outlying data points did not substantially impact model estimation.

The conditional model had a much larger average constant change value with much smaller proportional change values as compared to the unconditional model. This might imply that conditioning on these covariates helps explain variance in mathematics scores over time that would otherwise be captured by proportional change in the unconditional model, increasing the constant change rate, and leaving relatively less intraindividual variability to be captured within proportional change. In the conditional model, prior scores have essentially no detectable bearing on subsequent change for ELLs, although proportional change is still significant and positive for EPSs until Grades 2–3 and Grades 3–4 (albeit small in magnitude). This would suggest that even if ELLs and EPSs were equal on all covariates (as well as constant change mean and variance), prior knowledge continues to operate differently in predicting change between ELLs and EPSs.

(1b) To what extent does WM differ across ELLs and EPSs in uniquely predicting mathematics development patterns?

To answer Research Question 1b, we allowed the regressions of the MG-LCSM intercepts and slopes on WM to vary across groups while constraining all other regressions to equality across groups. When freed across groups, WM showed a significant effect for EPSs on both the intercept (EPSs $b = 0.158$, $SE = 0.008$, $b^* = 0.235$, $p < .001$) and the slope ($b = 0.018$, $SE = 0.007$, $b^* = 0.123$, $p = .006$). ELLs evidenced similar patterns (intercept: $b = 0.148$, $SE =$

0.033, $b^* = 0.221$, $p < .001$; slope: $b = 0.015$, $SE = 0.012$, $b^* = 0.103$, $p = .180$).[3] However, freeing the WM regressions did not improve the overall model fit based on the BIC (BIC increase from 172144.972 to 172161.892), suggesting that WM similarly predicted growth parameters across ELLs and EPSs (this is consistent with the similar point estimates).

Research Question 2

(2a) Are There Unobserved, Systematic, Differential Growth Trajectories (i.e., Mixtures) in Mathematics Development When the Significant Majority of ELLs Are Assessed in English (Grades 1–4)?

To address Research Question 2a, we chose a latent-basis growth model with freed residuals for GMM estimation. The latent-basis model fixes the first and last slope loadings to the desired time metric (e.g., 0 and 3 for four time points of equal interval) but leaves slope loadings for the remaining data points to be estimated. This produces a non-linear growth curve. After fitting multiple types of two- and three-class GMMs, only four estimable models remained for consideration (i.e., models that did not have non-positive-definite matrices). In evaluating mixture model fit, we considered the BIC as an index of relative model performance as well as entropy, which indicates the separability of classes. Higher entropy values suggest higher class distinctiveness, which is desirable when estimating latent classes.

BIC and entropy values from each mixture model (i.e., one-, two-, and three-class models) suggested a slight improvement in model fit from a one class (BIC = 8980.38) to two class model with freed parameter means (BIC = 8966.84, entropy = .83) and a two-class model with all parameters (means and variances) freed across classes (BIC = 8966.73, entropy = .95).

---

[3] Standardized coefficients (b^*) were calculated manually using the overall sample WJ-III NR *SD* from Table 1 (16.61) and the square roots of the latent intercept and latent slope variances (*SD*s) from Figure 1 (latent intercept *SD* = 11.13, latent slope *SD* = 2.43).

Balanced with entropy and BIC values, however, the classes did not appear qualitatively meaningful based on mixing proportions (≈.06–.09). Moreover, the only significant differences between mixture classes in the two-class models were the intercept values (approximately 60 in one class, ranging from approximately 36 and 42 in the second class depending on specification), the residual terms (when they were freed across time and class), and one growth factor loading in the latent-basis model (Time 2 loading approximately 0.50 in the second class with lower intercept versus approximately 1.4 in the other class). Otherwise, the trajectories closely modeled each other (e.g., similar latent slope means and Time 3 loading). The three-class model presented no advantage based on entropy or BIC values. The two-class mixture model was also sensitive to one outlier (a very low score) in Grade 4; two- and one-class models were more similar based on the BIC when removing the outlier. Model fit indices for each estimated model are presented in the Supplemental Materials (Table S3). All considered, we selected a single-class latent-basis growth model as the final model. The single-class model precluded the planned latent transition analysis (see Supplemental Materials).

The single-class model showed mixed fit χ^2 (*df*) = 14.519 (3), CFI = 0.991, TLI = 0.967, RMSEA = 0.113, SRMR = 0.114.[4] Parameter estimates for this unconditional latent-basis model are presented in Table 4 (Column 1). This model showed wide variability in residual variance over time; Time 2 (Grade 2) residual variance was more than twice as large as compared with Time 1 or Time 4. Growth rates also varied widely around the average rate of 13.98; 95% of average growth rates fell between approximately 8 and 20. Thus, by the end of Grade 4, students at the upper 95% bound of growth would be predicted to grow 60 points, whereas those at the

[4] CFI and TLI were calculated using an intercept-only model with residuals equal over time (χ^2 [*df*] = 1284.445 [11]).

lower 95% bound would be expected to grow 24 points. The difference in growth rate between the upper and lower 95% bounds (36) is larger than the lower and upper 95% bounds of the Grade 4 mathematics score distribution. The growth rate and intercept covariance was too imprecise to distinguish from 0, and the corresponding correlation was small in magnitude (r = 0.21), indicating that intercept values (Grade 1 mathematics scores) provide only minimal information about growth rate. Together, the wide variability in growth rate through Grade 4 and lack of detectable correlation with Grade 1 scores indicated that ELLs' Grade 4 scores emerged from a wide range of predicted (non-linear) trajectories that were imprecisely and minimally correlated with where students started in Grade 1. Higher or lower performance in Grade 1 had minimal bearing on students' expected performance by Grade 4.

(2b) Do Gains in ELP Across Kindergarten Predict Greater Mathematics Growth?

To address 2b, we introduced covariates into the single-class latent-basis growth model described above. We estimated separate models for *pre*LAS and EBRS gains. Table 4 (Columns 2 and 3) displays the growth parameters and fit indices for the two conditional growth models. Table 5 displays the effects of the covariates in each of the two conditional growth models (one with *pre*LAS gains and one with EBRS gains). *pre*LAS gains did not emerge as a significant predictor of later growth intercept (b^* = 0.064) or slope in mathematics (b^* = -0.078). In the EBRS model, the growth model intercept regression on EBRS change revealed a significant and positive relationship, indicating that students' growth in English early literacy predicted higher mathematics performance at the end of Grade 1 (b^* = 0.217). However, EBRS gains predicted a reduction in growth slope (b^* = - 0.217), suggesting that the advantage of higher Grade 1 (intercept) scores predicted by EBRS is somewhat offset by a lower growth rate through Grade 4.

(2c) Is ELP a Stronger Predictor of Mathematics Growth Patterns Among Students With Lower WM Capacity? If Mixture Classes Are Present, How Does This Interaction Operate Across Classes?

Table 5 displays the average effect of WM on mathematics growth parameters as well as WM's interaction with either *pre*LAS or EBRS gains. In neither model did WM significantly moderate the effects of *pre*LAS or EBRS gains, although WM showed a significant positive effect on intercept and a significant negative effect on slope in both conditional models. This suggests that irrespective of rate of language development across kindergarten, WM predicts higher Grade 1 mathematics scores but at the same time a reduction in growth through Grade 4.

Discussion

The imperative to equitably meet the educational needs of linguistically diverse students (Albers & Martinez, 2015; Robinson-Cimpian et al., 2016; Suhr et al., 2021; Sullivan, 2011), coupled with federal accountability mechanisms for achievement and ELP development (Every Student Succeeds Act, 2015), requires more investigation of mathematics growth patterns and their relationships with key predictors among ELLs. This study addressed differences between ELLs and EPSs in mathematics change patterns, potential heterogeneities in ELL mathematics development, and the interactions between WM and ELP in predicting mathematics development.

Our analysis showed that ELLs and EPSs changed in similar ways for the most part. The constant change rate was fixed to equality across groups, and the proportional change coefficients varied somewhat between ELLs and EPSs, although not to a degree that produces substantially different trajectories (on average). Our findings do not support our hypothesis that ELLs would grow more and then plateau more than EPS peers in later grades. Rather, EPSs showed slightly steeper growth initially, and their rate of change decreased more than ELLs (as evidenced by differences in proportional growth rates), although this was not substantial.

When including covariates, the MG-LCSM showed an increase in the constant growth rate and a decrease in proportional change estimates across groups, and the proportional change parameters were generally smaller among ELLs than EPSs as well. This potentially indicates that after accounting for covariates, the compounding intraindividual effects on change could operate somewhat differently across groups. For example, in the unconditional model, proportional change was more positive and larger for ELLs in some instances, but in the conditional model, proportional change was smaller in magnitude and negative in kindergarten and Grade 1. This would indicate that after accounting for covariates in the constant change rate, the intraindividual variability in change is weakened, particularly for ELLs in early elementary grades. Prior knowledge could carry less weight in shaping change for ELLs in early elementary (even when students are equated on all covariates, including kindergarten-entry mathematics knowledge), although growth becomes increasingly contingent on prior knowledge for all students in later grades. Nevertheless, these proportional effects in the conditional model remain small in magnitude and do not produce markedly different trajectories through Grade 4. Controlling for covariates, WM similarly and positively predicted growth parameters across ELLs and EPSs. Scoring higher on WM at the end of kindergarten was related to an increase in kindergarten mathematics scores as well as an increase in constant growth through Grade 4.

When assessed in English, ELLs showed a nonlinear trajectory from Grades 1–4. There was significant variance in growth trajectories, yet this did not vary systematically so as to suggest unobserved distributions of growth. In addition, the relationship between early English language skill growth and mathematics growth appeared to depend on the type of assessment administered (i.e., oral language vs. early literacy). English early literacy gains positively predicted Grade 1 mathematics performance but negatively predicted growth, whereas receptive and expressive oral language in English (via the *pre*LAS) did not significantly predict either level or growth rate in mathematics. The magnitude of these effects also differed, with EBRS showing larger standardized regression coefficients than the *pre*LAS.

These results have practical and research implications. Even after controlling for a variety of baseline characteristics, ELLs and EPSs changed in more similar ways than not, albeit at different levels of performance. The primary difference in change patterns came from the relationship of prior performance to subsequent change over time, but even this did not produce markedly different shapes of change. Including covariates in the MG-LCSM, however, led to (a) a significant shift in the estimated sources of change, (b) increases in the constant change parameter of more than three-fold, and (c) decreases in the proportional change effects (becoming negative in later grades). This suggests that accounting for the covariates may weigh the developmental process more heavily towards constant growth by decreasing the relation of prior scores to subsequent change (proportional effects). In addition, there was some variability in the size and precision of these proportional effects across groups, indicating that growth may compound less over time in early elementary for ELLs as compared with EPSs. Including covariates also removed initial performance (intercept) differences (see Figures 2 and 3).

For ELLs, constant change may be the primary driver of change particularly in earlier grades; prior achievement bears minimally (and imprecisely) on subsequent change until Grade 2 when these proportional effects become larger for both ELLs and EPSs. This is perhaps consistent with findings that prior knowledge becomes increasingly important across grade levels (Geary et al., 2017), yet it leaves questions regarding why prior skills are not detectably related to change for ELLs in earlier grades after conditioning on covariates. Extending these results to the context of MTSS, our findings underscore the need to understand the full developmental trajectory when targeting screening and prevention/intervention programs. Constant change is something that occurs as an additive function across the full developmental span (kindergarten–Grade 4, in our case), and understanding change between any two grade periods requires a sense of how students are expected to progress in the long-term because proportional change occurs relative to the underlying constant trajectory. We found that constant change was better estimated as equal across ELLs and EPSs, so that the long-term constant trajectories were similar, but this still leaves some differences in proportional change patterns. These results are also important to interpret in the context of the initial group differences of around 0.80 *SD*s and how the different distributions of mathematics scores for each group relate to these proportional change parameters in the unconditional model. Descriptively, ELLs performed lower at each wave than EPSs, meaning intraindividual change is occurring relative to different distributions of skills, offering one possible explanation of why proportional change may compound differently over time.

Despite the overall trajectories being similar on average, these intraindividual trends indicate that there is a strong need to understand *how* ELLs acquire, maintain, and compound their mathematics knowledge within the context of their classroom experiences. This is important irrespective of background characteristics that might further differentiate individuals. This would be key information to consider within the context of universal screening and monitoring students' growth seasonal or annual benchmarks (Albers & Martinez, 2015). For example, for states using the ACCESS ELP assessment (WIDA, 2022), WIDA's Can-Do descriptors can be used to better contextualize students' English academic content knowledge and skills with respect to their ELP development. These descriptors might help better describe how assessments represent students' skills, what expectations for growth in English mathematics might be reasonable, and what intraindividual developmental processes of mathematics and second language development might help school personnel understand students' change *relative to themselves* in the context of the overall developmental trajectory. Our current findings also suggest that the overall trajectory may mask important details about how each students' growth compounds over time relative to themselves. Understanding these sources of change may help facilitate understanding the strengths and needs of each student in terms of their mathematics content knowledge rather than focusing on aggregate trajectories that are characterized primarily by large achievement gaps relative to EPSs. Indeed, these types of normative and ipsative comparisons in the context of absolute (e.g., national norms) and relative norms (e.g., local norms, norms relative to ELLs) are exactly the kinds of data-based decisions that can lead to more equitable access to effective and appropriate instructional services (Albers & Maritnez, 2015; Robinson-Cimpian et al., 2016).

Our findings suggest that higher WM relates to sustained higher mathematics performance over time. This is similar for ELLs and EPSs alike based on Research Question 1b, thus corroborating prior work identifying WM as a positive predictor of mathematics performance and development (Geary et al., 2017; Miller-Cotto & Byrnes, 2020; Swanson et al., 2018, 2021). The domain-specificity of a digit span task (like numbers reversed) with regard to mathematics knowledge (Geary, 1993; Peng et al., 2018) may be in part one explanation for this consistent relation of WM to mathematics in our present study. Screening for skills such as WM could facilitate the identification of cognitive processing difficulties that are more agnostic to the specific content (Clemens et al., 2016), yet a more precise measurement of those skills would be necessary (WM subskills; Gathercole et al., 2004). That being said, incorporating skills such as WM into school-based screening may not be justified when domain-specific measures more strongly relate to mathematics (Clarke et al., 2018). In contrast to our hypotheses, we did not find evidence that WM moderated the relationship of EBRS or *pre*LAS growth to later mathematics growth factors.

Overall, WM's similar role across ELLs and EPSs and among only ELLs has broader developmental implications that suggest this central executive process is similarly predictive regardless of students' ELP or their rate of English acquisition. Consistent with our previous points, understanding students' language development and how that may relate to their content knowledge (and vice versa) is likely the most effective route to accurately targeted prevention and intervention. Considering the covariance of mathematics and WM (Geary et al., 2017; Miller-Cotto & Byrnes, 2020), these decisions would naturally include aspects related to WM, but it does not appear WM deserves differentially close attention among ELLs compared withEPSs or across ELLs' ELP growth rates.

The single class of growth curves among ELLs implies that a single, average trajectory was able to capture ELLs' mathematics growth. Remaining variation was captured in the intercept and slope variances, but this variation was not systematic, such as what might be found in mixture models (Hong & You, 2012; Jordan et al., 2006, 2007). However, we reemphasize that multiple data sources should be corroborated to establish the absolute, relative, and normative formulations of academic skill development among ELLs (Albers & Martinez, 2015). We show that *pre*LAS growth did not significantly relate to mathematics level or growth, although EBRS did. EBRS gains (i.e., early literacy skills) may be more indicative of students' understanding of grade-level instruction in reading (as well as other subjects that would be critical for mathematics development) than vocabulary alone. The negative slope effect of these gains in mathematics development, however, suggests that gains in early literacy alone may not relate to sustained mathematics growth over time. Strong emphases on language use in instruction among ELLs is important for developing mathematics skills (Doabler et al., 2016), although more research is needed on what facets of language best facilitate mathematics development. Recent work has addressed how different language measures predict mathematics (e.g., Chow & Ekholm, 2019), which could help inform what aspects of language to account for in assessment, instruction, and intervention.

These findings confirm that attention to both mathematics content knowledge and early English literacy skills are key considerations in promoting the positive mathematics growth of ELLs. Although WM independently predicts mathematics trends, which is important to consider in a broader developmental sense, specific targeting of domain-general processes such as this are unlikely to yield favorable balance of costs-benefits (Clarke et al., 2018), particularly when one of the most important inequities to address in ELLs' academic development is exposure to effective grade-level instruction exposure comparable with English monolingual peers (Robinson-Cimpian et al., 2016).

Limitations

A number of limitations are important to note. First, we did not account for individually-varying times of observations in our analysis. Some alternative specifications may better capture this (McArdle & Nesselroade, 2014; Sterba, 2014); however, a continuous time framework was not feasible in our current models. Moreover, LCSM constant and proportional change parameters are frequently very strongly correlated (Jacobucci et al., 2019); our models showed this same limitation ($r > .95$). This suggests that although proportional and constant change are important in practical terms, they may not be as clearly disentangled as LCSMs suggest. Using Spring 2011 WM (instead of Fall 2010) raises temporality questions in Research Question 2, given that we used WM as a moderator of Fall-Spring ELP growth, and this is also relevant to Research Question 1. Last, WM has multiple subcomponents (Gathercole et al., 2004) and is itself part of a larger constellation of cognitive abilities, although WM capacity is a key feature underlying a variety of cognitive tasks (Kovacs & Conway, 2016). Only one WM measure was available in the ECLS-K: 2011 data, so it is yet to be determined based on the current data the extent to which WM uniquely predicts math achievement when considering the larger scope of cognitive abilities (Dombrowski et al., 2018). This is a limitation that extends beyond the scope of this study, although our current WM findings are robust to the inclusion of a variety of other cognitive, behavioral, and achievement measures, suggesting that WM nonetheless possesses a unique relation to mathematics (even when accounting for prior mathematics performance). Measurement of language faces a similar issue in that there are a limited number of language measures available in the ECLS-K: 2011.

These findings should be interpreted with consideration of the practical limitation of defining “ELL” using this dataset. It is important to remain conscientious of the assumptions embedded within ELL classification schemes, given how these definitions align with practice and shift the narrative of performance among linguistically diverse students (Kieffer & Thompson, 2018). Our results are admittedly subject to this exact problem given our definition of ELLs, although we hoped to strike a balance between the specificity of the ECLS-K: 2011 definition for assessment purposes and sensitivity of parent-reported non-English language speaking.

Despite these limitations, however, we also see many strengths in using large-scale national data to answers questions such as those discussed here. Of course, large-scale national data in many ways lose the focus on context that is extremely important to considering mathematics growth among ELLs and EPSs. However, large-scale data possess unique capabilities to make stronger generalizations (under certain circumstances) that are not attainable with smaller samples. This has to be balanced with the depth of measurement, which is inherently more limited in this type of data given the scope of the overall study. Thus, although our measures of ELP, ELL, and WM may be less desirable than more thorough batteries of language and cognitive functioning, these data still possess unique strengths in terms of the scope and breadth of data collected.

Conclusion

This study supplements existing findings that highlight persistent achievement gaps between ELLs and non-ELLs. The current work adds unique perspective to this literature by using techniques to attempt to uniquely partition sources of mathematics development trends and variability, which has not been sufficiently documented in previous studies among ELLs with limited ELP. Understanding sources of change and systematic variability in change helps answer practical questions about who changes, when, and what predicts such change. These factors are central to data-based decision-making within MTSS and making informed, equitable, and effective data-based decisions. The current results help better understand how ELLs' grow over time and the sources of variability in growth, which will help inform the equitable targeting of resources within MTSS to meet ELLs' mathematics learning needs.

References

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

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

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338–363. doi: 10.1037/1082-989X.8.3.338

Cáncer, P. F., Estrada, E., Ollero, M. J. F., & Ferrer, E. (2021). Dynamical properties and conceptual interpretation of latent change score models. *Frontiers in Psychology*, *12*, 696419. https://doi.org/10.3389/fpsyg.2021.696419

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.2018.02.011

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*(4), 463–502. doi: 10.1080/00273171.2013.787870

Clarke, B., Shanley, L., Kosty, D., Baker, S. K., Cary, M. S., Fien, H., & Smolkowski, K. (2018). Investigating the incremental validity of cognitive variables in early mathematics screening. *School Psychology Quarterly, 33*(2), 264–271. doi: 10.1037/spq0000214

Clemens, N. H., Keller-Margulis, M. A., Scholten, T. S., & Yoon, M. (2016). Screening assessment within a multi-tiered system of support: Current practices, advances, and next steps. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of multi-tiered systems of support* (2nd ed., pp. 187–214). Springer.

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

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

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). Hierarchical exploratory factor analyses of the Woodcock-Johnson IV Full Test Battery: Implications for CHC application in school psychology. *School Psychology Quarterly*, *33*(2), 235–250. https://doi.org/10.1037/spq0000221

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

Every Student Succeeds Act of 2015, 20 U.S.C 70. § 6301 *et seq* (2015).

Friso-van den Bos, I., Kroesbergen, E. H., & van Luit, J. E. (2014). Number sense in kindergarten children: Factor structure and working memory predictors. *Learning and Individual Differences*, *33*, 23–29. http://dx.doi.org/10.1016/j.lindif.2014.05.003

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of WM from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. doi: 10.1037/0012-1649.40.2.177

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin*, *114*(2), 345–362. https://doi.org/10.1037/0033-2909.114.2.345

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*(5), 680–693.

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

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., Serang, S., & Grimm, K. J. (2019). A short note on interpretation in the dual change score model. *Structural Equation Modeling*, *26*(6), 924–930. doi: 10.1080/10705511.2019.1619457

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*(1), 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*(1), 153–175. doi: 10.1111/j.1467-8624.2006.00862.x

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance for education policy. *Peabody Journal of Education*, *77*(4), 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*(2), 457–467. doi: 10.1037/0012-1649.44.2.457

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

Kellam, S. G., Koretz, D., & Mościcki, E. K. (1999). Core elements of developmental epidemiologically based prevention research. *American Journal of Community Psychology*, *27*(4), 463–482.

Kieffer, M. J., & Thompson, K. D. (2018). Hidden progress of multilingual students on NAEP. *Educational Researcher*, *47*(6), 391–398. doi: 10.3102/0013189X18777740

Kovacs, K., & Conway, A. R. A. (2016). Process overlap theory: A unified account of the general factor of intelligence. *Psychological Inquiry*, *27*(3), 151-177. https://doi.org/10.1080/1047840X.2016.1153946

Lumley, T. (2020). *survey: Analysis of complex survey samples*. R package version 3.34.

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

Miller-Cotto, D., & Byrnes, J. P. (2020). What's the best way to characterize the relationship? between working memory and achievement? An initial examination of competing theories. *Journal of Educational Psychology*, *112*(5), 1074–1084. doi: 10.1037/edu0000395

Múthen, B., & Múthen, L. (2012-2020). *Mplus version 8.5*. [Computer software].

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

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling*, *21*(3), 439–454.

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology*, *108*(4), 455–473. https://doi.org/10.1037/edu0000079

Peng, P., Wang, C., & Namkung, J. (2018). Understanding the cognition related to mathematics difficulties: A meta-analysis on the cognitive deficit profiles and the bottleneck theory. *Review of Educational Research*, *88*(3), 434-476. https://doi.org/10.3102/0034654317753350

Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin*, *146*(7), 595–634. https://doi.org/10.1037/bul0000231

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*(1), 103–113.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual differences and cognitive approaches. *Learning and Individual Differences*, *20*, 110–122. http://dx.doi.org/10.1016/j.lindif.2009.10.005

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*(6), 565–576. doi: 10.1177/0165025409343765

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*(4), 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 Brain Sciences, 3*(1), 129–137.

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

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

Sterba, S.K. (2014). Fitting nonlinear latent growth curve models with individually-varying time points. *Structural Equation Modeling*, *21*, 630–647. https://doi.org/10.1080/10705511.2014.919828

Suhr, M. P., Nese, J. F. T., & Alonzo, J. (2021). Parallel reading and mathematics growth for English learners: Does timing of reclassification matter? *Journal of School Psychology, 85*, 94–112.

Sullivan, A. L. (2011). Disproportionality in special education identification and placement of English language learners. *Exceptional Children*, *77*(3), 317–334.

Swanson, H. L., Arizmendi, G. D., & Li, J-. T. (2021). Working memory growth predicts mathematical problem-solving growth among emergent bilingual children. *Journal of Experimental Child Psychology*, *201*, 104988. https://doi.org/10.1016/j.jecp.2020.104988

Swanson, H. L., Kong, J., & Petcu, S. (2018). Math difficulties and WM growth in English language learner children: Does bilingual proficiency play a significant role? *Language,* *Speech, and Hearing Services in Schools*, *49*, 379–394. https://doi.org/10.1044/2018_LSHSS-17-0098

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–288. https://doi.org/10.1207/s15516709cog1202_4

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

Tourangeau, K., Nord, C., Le, T., Wallner-Allen, K., Vaden-Kiernan, N., Blaker, L...Mulligan, G. M. (2015). *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*. 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*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

U.S. Department of Education. (2016). Non-regulatory guidance: English learners and Title III of the Elementary and Secondary Education Act (ESEA), as amended by the Every Student Succeeds Act (ESSA). Author. https://www2.ed.gov/policy/elsec/leg/essa/essatitleiiiguidenglishlearners92016.pdf

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). *mice*: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67.

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(2)*,* 227–244. doi: 10.1016/j.jecp.2013.02.002

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

WIDA. (2022). *ACCESS for ELLs*. Author. https://wida.wisc.edu/assess/access

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

Woodcock, R. W, & McGrew, K. S. (2001). *Woodcock-Johnson Test of Cognitive Abilities – III*. Riverside Publishing.

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

Table 1

EPS and ELL Weighted Sample Descriptive Statistics

	M			*SD*			Minimum			Maximum			Proportion Missing		
	EPS	ELL	All	EPS	ELL	All	EPS	ELL	All	EPS	ELL	All	EPS	ELL	All
EBRS – Fall K	13.95	8.05	13.60	4.07	4.63	4.30	0.00	0.00	0.00	20.00	20.00	20.00	.000	.000	.000
EBRS – Sp. K	17.58	13.65	17.38	2.47	4.21	2.76	0.00	0.00	0.00	20.00	20.00	20.00	.000	.000	.000
IRT Mathematics – Fall K	35.88	23.73	35.15	11.34	7.61	11.52	10.25	9.79	10.25	139.16	48.54	139.16	.000	.000	.000
IRT Mathematics – Sp. K	49.82	38.24	49.12	12.17	10.60	12.38	13.58	14.13	13.58	98.29	76.11	98.29	.000	.000	.000
IRT Mathematics – Sp. Gr. 1	74.33	58.93	73.40	16.69	13.90	16.93	21.91	25.88	21.91	131.82	103.40	131.82	.001	.000	.001
IRT Mathematics – Sp. Gr. 2	90.93	78.10	90.16	15.68	16.05	15.99	14.79	23.45	14.79	143.97	115.43	143.97	.001	.000	.001
IRT Mathematics – Sp. Gr. 3	103.18	92.66	102.55	14.96	15.17	14.96	40.25	43.95	40.25	144.25	123.91	144.25	.002	.003	.002
IRT Mathematics – Sp. Gr. 4	110.36	100.87	109.79	14.69	14.77	14.86	25.22	28.16	25.22	139.06	130.95	139.06	.002	.003	.002
WJ - III NR–	96.44	85.22	95.77	16.40	16.46	16.61	41.00	51.00	41.00	157.00	134.00	157.00	.000	.000	.000

Spring K															
DCCS – Fall K	14.58	12.09	14.43	2.99	4.12	3.12	0.00	0.00	0.00	18.00	18.00	18.00	.000	.000	.000
CBQ Attentional Focus – Fall K	4.79	4.39	4.77	1.27	1.33	1.28	1.00	1.00	1.00	7.00	7.00	7.00	.035	.063	.035
CBQ Inhibitory Control – Fall K	5.00	4.73	4.98	1.24	1.31	1.25	1.00	1.00	1.00	7.00	7.00	7.00	.038	.056	.038
Age at Kindergarten Entry	66.44	65.02	66.35	4.44	4.84	4.48	39.83	39.10	39.83	86.87	81.07	86.87	.002	.003	.002
Socioeconomic Status	-0.09	-0.84	-0.14	0.75	0.54	0.76	-2.07	-2.33	- 2.33	2.60	1.61	2.60	.002	.003	.002
*pre*LAS Score – Fall K	28.69	13.56	27.78	2.32	7.59	4.62	5.00	0.00	0.00	30.00	22.00	30.00	.000	.000	.000
*pre*LAS Score – Sp. K	29.32	22.04	28.91	1.73	6.74	2.92	0.00	0.00	0.00	30.00	30.00	30.00	.000	.000	.000
Unweighted *N* (EPS/**ELL**/*All*)	4,711/**303**/*5,014*														
Weighted *N* (EPS/**ELL**/*All*)	3,226,117/**205,946**/*3,432,073*														

Note. Dimensional Change Card Sort, NR = Numbers Reversed, CBQ = Children's Behavior Questionnaire, EBRS = English Basic Reading Skills, IRT = Item response theory. Descriptive statistics were adjusted for the complex survey design using the R package *survey* (Lumley, 2020).

Table 2

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	.55	.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
Hawaiian/Native Pacific Islander	.003	.01		.004	.000	.000
American Indian/Alaska 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. Descriptive statistics were adjusted for the complex survey design using the R package *survey* (Lumley, 2020). *$p < .05$, **$p < .01$, ***$p < .001$

Table 4

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

Predictor	Regression Coefficient (SE)	
	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 Dimensional Card Sort	0.190 (0.045)***	0.099 (0.029)**
Fall K Attentional Focus	0.639 (0.152)***	0.234 (0.078)**
Fall K Inhibitory Control	0.277 (0.132)*	-0.094 (0.066)
Fall K EBRS	0.081 (0.046)	-0.018 (0.017)
Fall K Full-Day	0.900 (0.438)*	-0.569 (0.173)**
Parental Pre-K Care	-0.192 (0.357)	0.115 (0.150)
Age at Kindergarten Entry	0.171 (0.039)***	-0.092 (0.016)***
Not First-Time Kindergartener	1.121 (0.648)	-2.509 (0.406)***
Socioeconomic Status	0.926 (0.196)***	0.467 (0.106)***
Non-Hispanic	0.110 (0.386)	-0.127 (0.172)
Female	-0.984 (0.219)***	-1.239 (0.138)***
Fall K Mathematics IRT	0.678 (0.019)***	-0.006 (0.017)
R^2	.81/**.70**	.21/**.19**
RMSEA	0.048	
SRMR	0.062	

BIC	172144.972
χ^2 (*df*)	822.864 (120)

Note. Race/ethnicity indicators collapsed into a single indicator of non-Hispanic status. All predictors centered at their grand means (socioeconomic status uncentered as it is already standardized to have a mean of 0 and *SD* of 1). Bolded coefficients represent values for ELLs. K = Kindergarten. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Table 5

Latent-Basis Growth Curve Parameters

Parameter	Unconditional Model (*SE*)	*pre*LAS Conditional Model Estimate (*SE*)	EBRS Conditional Model Estimate (*SE*)
Latent Intercept	58.908 (1.096)***	56.122 (1.342)***	57.717 (1.605)***
Latent Intercept Variance	175.325	92.226	84.782
Latent Slope	13.982 (0.223)***	14.967 (0.611)***	14.654 (0.657)***
Latent Slope Variance	8.553	6.849	6.393
Intercept-Slope Covariance	-8.187 (4.683)^	-5.250 (3.565)^	-3.564 (3.165)^
Latent Slope Factor Loadings			
Sp. Gr. 1	Fixed at 0	Fixed at 0	Fixed at 0
Sp. Gr. 2	1.379 (0.051)***	1.377 (0.050)***	1.375 (0.051)***
Sp. Gr. 3	2.414 (0.031)***	2.415 (0.031)***	2.414 (0.031)***
Sp. Gr. 4	Fixed at 3	Fixed at 3	Fixed at 3
Residual Variances			

Sp. Gr. 1	24.461	27.258	29.533
Sp. Gr. 2	53.700	52.530	51.859
Sp. Gr. 3	37.682	37.362	37.289
Sp. Gr. 4	22.143	22.897	22.960
Fit Indices			
R^2 Latent Intercept	--	.471	.511
R^2 Latent Slope	--	.159	.190
RMSEA	0.112	0.053	0.056
SRMR	0.114	0.055	0.056
BIC	8980.380	8943.506	8926.687
χ^2 (*df*)	14.344 (3)	60.840 (33)	64.159 (33)

Note. All predictors centered at ELL-specific means in conditional models. *p*-values not provided for fit indices and residual variances. ^$p > .05$ ***$p < .001$

Table 6

Covariate Effects in Latent-Basis Growth Curve Model

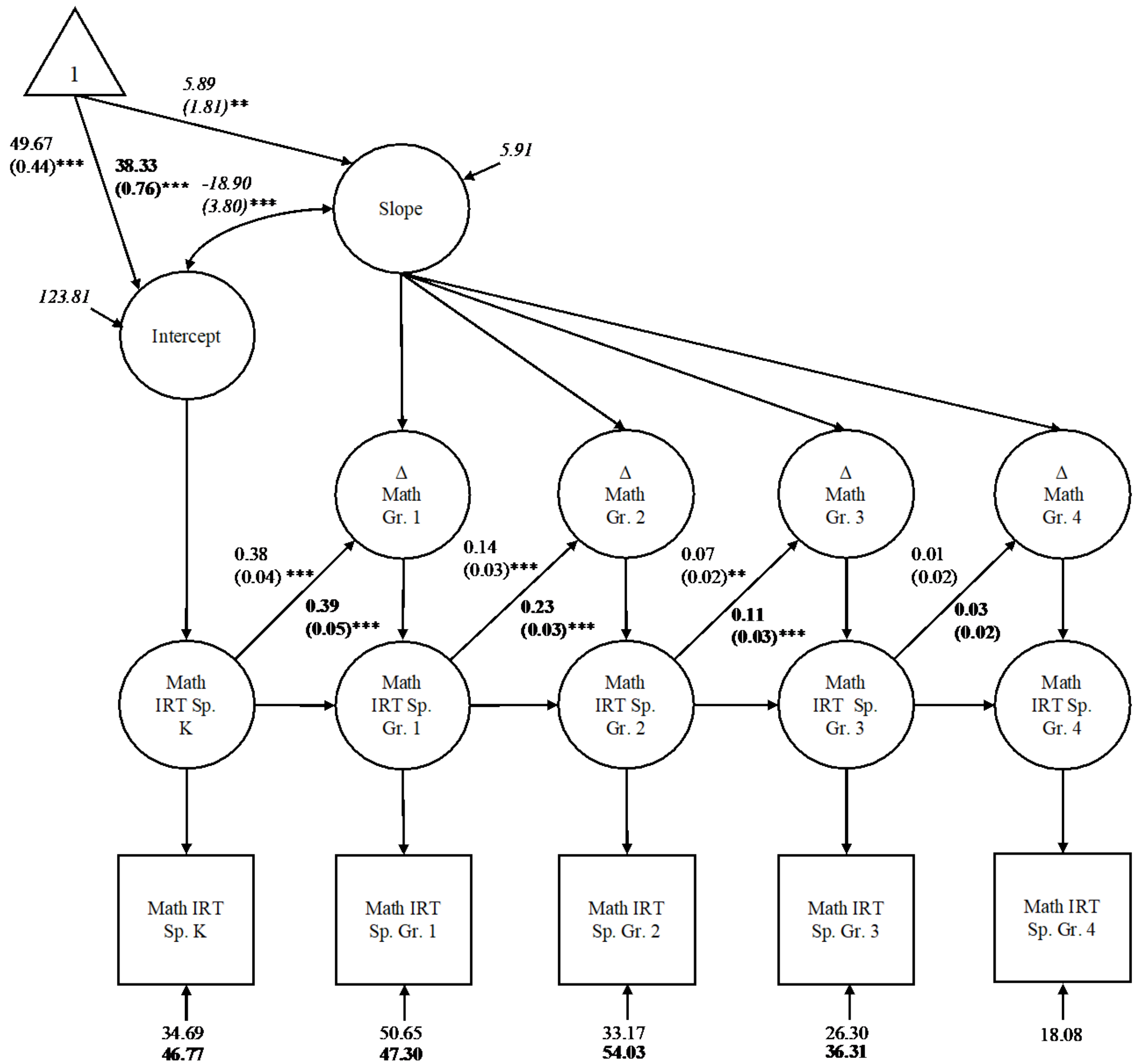
Predictor	Regression Coefficients (SE)	
	Growth Intercept	Growth Slope
***pre*LAS Model**		
(Intercept)	56.122 (1.346)***	14.967 (0.611)***
Sp. K. Working Memory	0.239 (0.059)***	-0.039 (0.013)**
*pre*LAS Gain Score	0.115 (0.100)	-0.031 (0.027)
Sp. K Working Memory X *pre*LAS Gain Score	0.002 (0.006)	0.001 (0.002)
Fall K Cognitive Flexibility	0.537 (0.154)**	0.009 (0.060)
Fall K Attentional Focus	1.516 (0.705)*	0.548 (0.195)**
Fall K Inhibitory Control	0.201 (0.845)	-0.479 (0.209)*
Fall K Full-Day	2.784 (1.451)	-0.255 (0.581)
Parental Pre-K Care	-0.092 (1.631)	-0.414 (0.660)
Age of Kindergarten Entry	0.081 (0.181)	-0.156 (0.052)**
Not First-Time Kindergartener	4.270 (2.859)	-2.318 (0.941)*
Socioeconomic Status	2.106 (1.233)	-0.832 (0.415)
Non-Hispanic or Asian	0.571 (2.963)	0.419 (1.923)
Asian	5.748 (3.140)	1.926 (0.996)
Female	-0.479 (1.196)	-1.201 (0.404)**
Fall K Mathematics IRT	0.525 (0.099)	0.031 (0.034)
EBRS Model		

(Intercept)	57.717 (1.605)***	14.654 (0.657)***
Sp. K. Working Memory	0.227 (0.052)***	-0.038 (0.012)**
EBRS Gain Score	0.576 (0.161)***	-0.104 (0.041)*
Sp. K Working memory X EBRS Gain Score	-0.005 (0.007)	0.002 (0.003)
Fall K Cognitive Flexibility	0.486 (0.160)**	0.019 (0.060)
Fall K Attentional Focus	1.581 (0.682)*	0.505 (0.211)*
Fall K Inhibitory Control	0.060 (0.843)	-0.437 (0.207)*
Fall K Full-Day	1.325 (1.546)	0.021 (0.616)
Parental Pre-K Care	-0.292 (1.438)	-0.379 (0.623)
Age of Kindergarten Entry	0.125 (0.163)	-0.165 (0.054)**
Not First-Time Kindergartener	5.211 (2.632)*	-2.479 (0.900)**
Socioeconomic Status	2.564 (1.168)*	-0.893 (0.407)*
Non-Hispanic or Asian	0.271 (3.057)	0.461 (1.891)
Asian	6.316 (3.000)*	1.788 (0.953)
Female	-1.024 (1.172)	-1.061 (0.415)*
Fall K Mathematics IRT	0.620 (0.097)***	0.019 (0.034)

Note. All predictors centered at ELL means. *$p < .05$, ** $p < .01$, ***$p < .001$

Figure 1

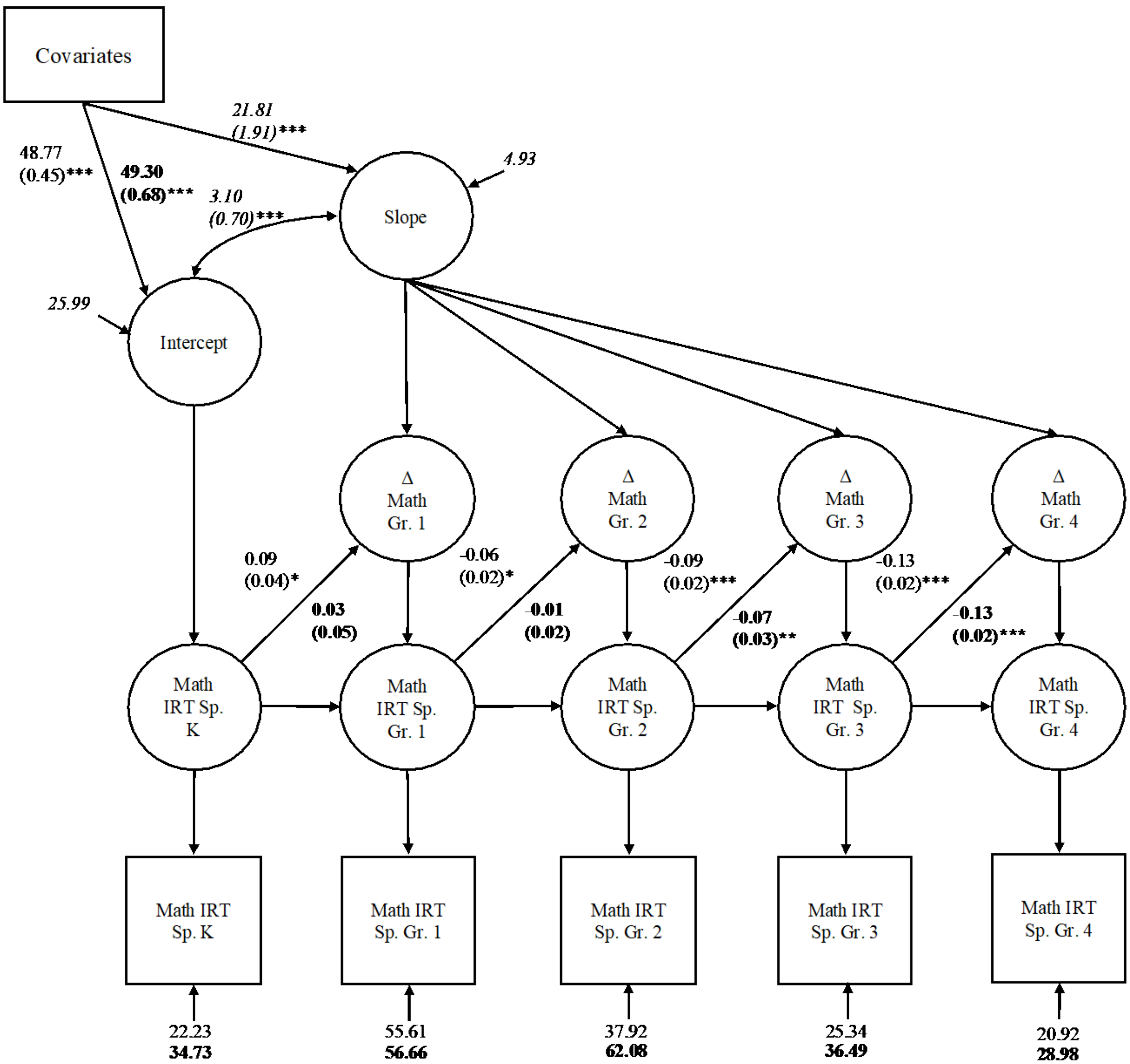
Unconditional MG-LCSM



Note. Unlabeled paths fixed to 1. Paths for ELLs are bolded. Italicized coefficients denote it is fixed across groups. Standard errors in parentheses. *p < .05, **p < .01, ***p <.001.

Figure 2

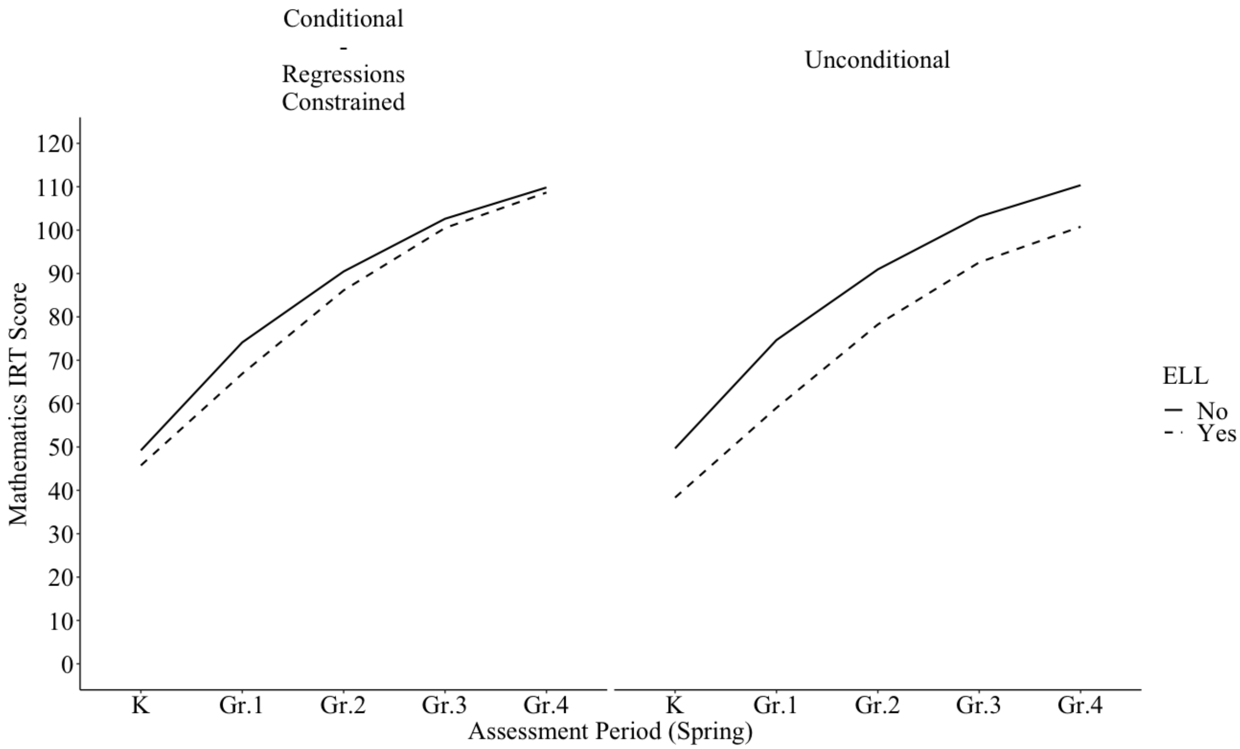
Conditional MG-LCSM Estimates with Regression Parameters Constrained to Equality Across Groups



Note. Unlabeled paths fixed to 1. Paths for ELLs are bolded. Italicized coefficients denote it is fixed across groups. Standard errors in parentheses. **p* < .05, ***p* < .01, ****p* <.001.

Figure 3

Estimated Latent Trajectories for Each MG-LCSM.



Note. Conditional intercept and slope estimates calculated using the means of each covariate within each group (including dummy codes) except Fall Kindergarten mathematics score, which was fixed at 0 across groups. Change scores in the unconditional trajectories calculated using the observed means for each group at each time point. Conditional trajectories were constructed by calculating change scores using observed means at each time point for each group in combination with the conditional latent slope calculated with covariates fixed to the values described above. These change scores were then added to the conditional intercepts for each group (calculated as described above) to create the trajectories for each group. Figure produced in *ggplot2* (Wickham, 2016).

Online Supplemental Materials for Modeling Associations of English Proficiency and Working Memory with Mathematics Growth

These Supplemental Materials provide additional information regarding (a) our criteria for determining ELL status, (b) information on the design of the study, (c) variables in our multiple imputation procedure, (d) Bayesian Information Criterion (BIC) values for unconditional latent change score models, (e) simulated trajectories of unconditional multiple-group latent change model (MG-LCSM), and (f) results of growth mixture modeling (GMM). Information on ELL classification is presented in Table S1, Figures S1, and Figure S2. Imputation information is presented in Table S2. Figure S3 displays MG-LCSM trajectories that are simulated from the model results displayed in the main text. GMM results are presented in Table S3. We also provide more description of these supplemental elements below.

The ECLS-K: 2011 used a three-stage complex sampling design with unequal selection probability for some students (Tourangeau et al., 2018). We used sampling design elements (weight variable [W8C18P_8T180], primary sampling unit [PSU; variable W8C18P_8T18PSU], and strata [W8C18P_8T18STR]) in all descriptive and inferential analyses. We subset the data for the ELL-only analyses (Research Question 2) using Mplus's SUBPOPULATION command.

The first table, Table S1, provides a cross-tabulation of proportions of students scoring below four different English language proficiency (ELP) cutoffs that we defined in the data using *pre*LAS scores (i.e., English proficient, Marginal English Proficiency, Limited English Proficiency, Very Limited English Proficiency) across parent-report primary home language in kindergarten (English or not English). We included students with a nonzero value on the base year sampling weight W2C_2P_2TZ0 and nonmissing Fall kindergarten *pre*LAS data and parent-reported home language (n = 12,425). The Pearson χ^2 test with Rao and Scott adjustment in the *survey* package of R (Lumley, 2020) showed that the proportions of each category between the two groups were significantly different (F = 530.77, numerator df = 2.77, denominator df = 443.16, p < .05). Figure S1 provides a graph of the raw *pre*LAS data and unweighted boxplots in this base year sample by kindergarten-year primary home language status (English or not English). As shown in the graph of these raw data, *pre*LAS scores among primarily English-speaking students were sparse below values of 23, whereas *pre*LAS scores were more evenly distributed among students with a non-English primary home language below a value of 23. We present the survey-weighted boxplots of the *pre*LAS scores in Figure S2. Both figures indicate similar results. The ECLS-K: 2011 used a *pre*LAS score of < 16 out of 30 as a cut score for determining eligibility for Spanish cognitive assessment (for Spanish-speaking students; see Tourangeau et al., 2018, for routing procedures). Our definition, then, is less restrictive than the ECLS-K: 2011 definition and includes a wider range of English language proficiency skills. Our method also attempts to distinguish from the typical range of *pre*LAS scores among primarily English- and non-English speaking students (i.e., scores between 23 and 30) and scores outside that range that may be indicative of more limited *pre*LAS skills among non-English speakers (i.e., scores below 23). Our intention in capturing this range of ELP was to avoid an over-restrictive criterion, such as the ECLS-K: 2011 definition, but also to avoid overinclusion of students with non-English primary home language who may be English-proficient into our definition of ELL. Our criterion does not necessarily align with exactly who took assessments in Spanish through the spring of first grade, however. Parent-reported home language may have been English (or cannot choose) yet the child scored below the *pre*LAS cutoff and still qualified to take the Spanish assessment. We included these students as English-proficient students given the reported primary home language of English. Only eight students in our final analytic sample were "misclassified" in this manner.

Table S2 provides a full list of the auxiliary variables included in our predictive mean matching imputation (PMM) procedure. For PMM, we included auxiliary variables (Enders, 2010), as well as the survey design elements (sampling weight and k -1 dummy indicators for the strata and PSUs) in the imputation procedure; Table S2 shows all variables included in PMM. We imputed missingness on all primary and auxiliary variables and using 20 datasets and five iterations each. We then exported imputed datasets to M*plus* Version 8.5 (Muthén & Muthén, 2017) for analysis.

Simulated Trajectories from Unconditional Multiple-Group Latent Change Score Model

In Figure S3, we present trajectories that are simulated from the model parameters displayed in Figure 1 of the main text. These trajectories help visualize what this model is predicting for individual growth trajectories. However, these results are not the same as estimated growth trajectories based on the actual data (these plots are not available from Mplus when using Mplus on a Mac). Rather, the simulations display what the observed data would look like if they were generated from a multiple-group latent change model with the model parameters provided in Figure 1. We used the R package *lavaan* (Rosseel, 2012) to simulate data from the Figure 1 model parameters. We created two *lavaan* models (one for ELLs and one for EPSs) matching exactly what we estimated in the multiple-group Mplus model, and we fixed all the parameters of the *lavaan* models to the exact values estimated from Mplus. Then, we used the simulateData() syntax in *lavaan* to generate 100 observations for each model (i.e., 100 observations for ELLs and 100 observations for EPSs).

As Figure S3 shows, the generated data for each group differ in a few ways. First, ELLs appear to show slightly more variability in their trajectories in addition to their time-specific scores (this is consistent with the relatively higher residual variances in the model among ELLs; see Figure 1). In addition, ELLs, on average, do not decelerate to the same extent as EPSs, which is consistent with the model parameters and the average trajectories shown in Figure 3 in the main text.

Information on Research Question 2: Growth Mixture Modeling

Addressing Research Question 2 consisted of multiple exploratory stages using growth mixture modeling. As noted in the main text, the first step is to determine if mixture classes are detectable. If mixture classes are present, the second step of the modeling process is to include covariates using the three-step method (Asparouhov & Múthen, 2014). The third step is testing whether students remain in their mixture class over time (Chow et al., 2013; Kaplan, 2008) and whether the covariates predict class membership change. Mplus's three-step procedure handles only one set of latent mixture classes, however, so we planned to use a latent transition analysis (LTA) with the three-step procedure proposed by Nylund-Gibson et al. (2014). Specifically, we planned to conduct separate latent profile analyses (LPAs) for each of the four mathematics measurements utilized in the GMM (Spring Grade 1 to Spring Grade 4) and extract profile memberships at each time point to replicate class memberships modeled in the GMM. We then planned to use LTA by regressing predicted LPA class membership to membership at the previous time point and including predictors (however, as discussed later, we selected a one-class GMM, precluding the use of the LTA). We first fit one-, two-, and three-class unconditional GMMs 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 (Ram & Grimm, 2009). We considered multiple practical (e.g., mixing proportions) and quantitative factors (relative model fit) to determine the appropriate number of classes (Bauer & Curran, 2003; Kaplan, 2002).

Figure S4 provides a sample of 30 ELLs' observed mathematics growth trajectories from Grades 1–4. Plotting such figures can help aid in the detection of mixture classes by providing qualitative information about the types of trajectories that may characterize mixture classes (Kaplan, 2002). Table S3 provides model fit indices for each growth mixture model in addition to the models with one outlier removed (all of this student's data were removed; results were similar when simply recoding the outlying value to a more central value).

References

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

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338–363. doi: 10.1037/1082-989X.8.3.338

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*(4), 463–502. doi: 10.1080/00273171.2013.787870

Enders, C. K. (2010). *Applied missing data analysis.* Guilford.

Kaplan, D. (2002). Methodological advances in the analysis of individual growth with relevance for education policy. *Peabody Journal of Education*, *77*(4), 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*(2), 457–467. doi: 10.1037/0012-1649.44.2.457

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

Nylund-Gibson, K., Grimm, R., Quirk, M., & Furlong, M. (2014). A latent transition mixture model using the three-step specification. *Structural Equation Modeling*, *21*(3), 439–454

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*(6), 565–576. doi: 10.1177/0165025409343765

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://www.jstatsoft.org/v48/i02/.

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*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics.

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

Table S1

Percentages of Students in Language Proficiency Categories

	Speaks English as Primary Home Language	
Language Proficiency Level (*pre*LAS score)	Yes	No
English Proficient (>= 28)	85.91% (94.05%)	33.05% (45.05%)
Marginal English Proficiency (<28)	12.08% (12.13%)	34.55% (34.54%)
Limited English Proficiency (<23)	1.04% (1.06%)	12.22% (12.15%)
Very Limited English Proficiency (<17)	0.88% (0.89%)	20.18% (18.25%)

Note. Unweighted percentages in parentheses.

Table S2

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 Spring 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 [race or no race specified combined] 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. Students who were retained in kindergarten in 2011-2012 received a slightly modified version of the CBQ, which was recorded as a separate variable in the dataset (Tourangeau et al., 2018). The CBQ was scaled identically and was combined with the students who advanced to first grade for the purposes of imputation. Students' grade year reflects their year in the study, not necessarily their enrolled grade (if they did not advance with peers; Tourangeau et al., 2018). CBQ = Children's Behavior Questionnaire; TMCQ = The Middle Childhood Questionnaire; WJ = Woodcock-Johnson; IRT = Item Response Theory.

Table S3
Unconditional MG-LCSM BIC Values

	BIC	ΔBIC from Model 1
Model 1 – Single Group Model		
Single group model (residuals freed over time, proportional change constrained over time)	179995.635	--
Model 2 – Multiple Group Model		
All change parameters constrained across groups including intercept (residuals and proportional change constrained across time and groups)	180989.506	993.871
Model 3 – Multiple Group Model		
Intercept freed (residuals and proportional change constrained across time and groups)	180807.937	812.302
Model 4 – Multiple Group Model		
Intercept + proportional change + residuals freed across time and groups	**178922.785**	**-1072.85**
Model 5 – Multiple Group Model		
Intercept + slope + residuals + variance/covariances freed + proportional change freed across time (but not group)	178936.862	-1058.773
Model 6 – Multiple Group Model		
Intercept + slope + residuals + variances/ covariances + proportional change freed across time and group	178932.065	-1063.570

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

Table S4

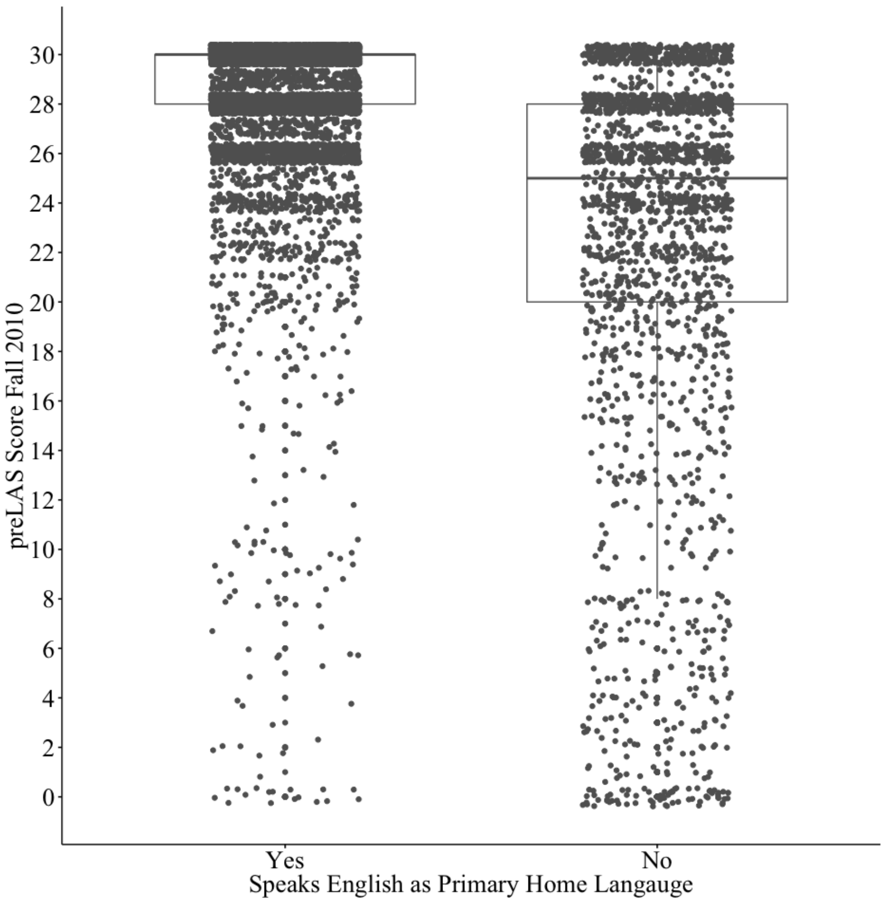
Relative Fit Indices of Growth Mixture Models

	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.

Figure S1

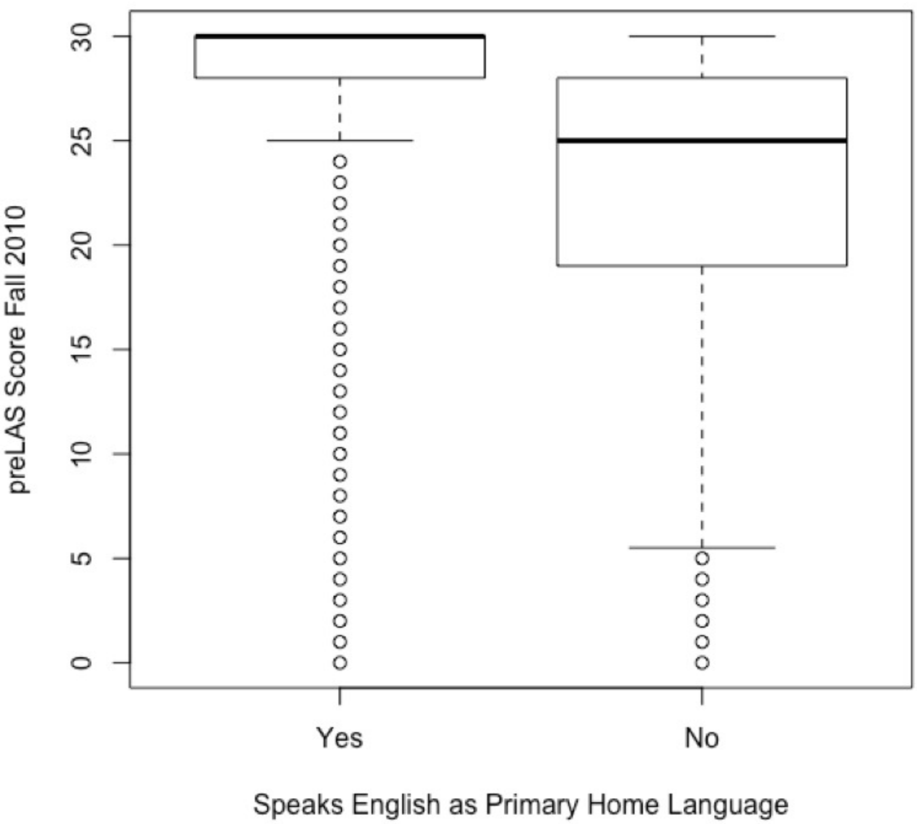
Scatter and Boxplots of Fall Kindergarten preLAS Scores by Parent-Reported Primary Home Language



Note. Parents who reported speaking multiple languages or could not choose are included in "Yes." The total unweighted sample size for this graph is 12,425. Boxplots are not adjusted for complex survey elements. Plot produced in *ggplot2* (Wickham, 2016).

Figure S2

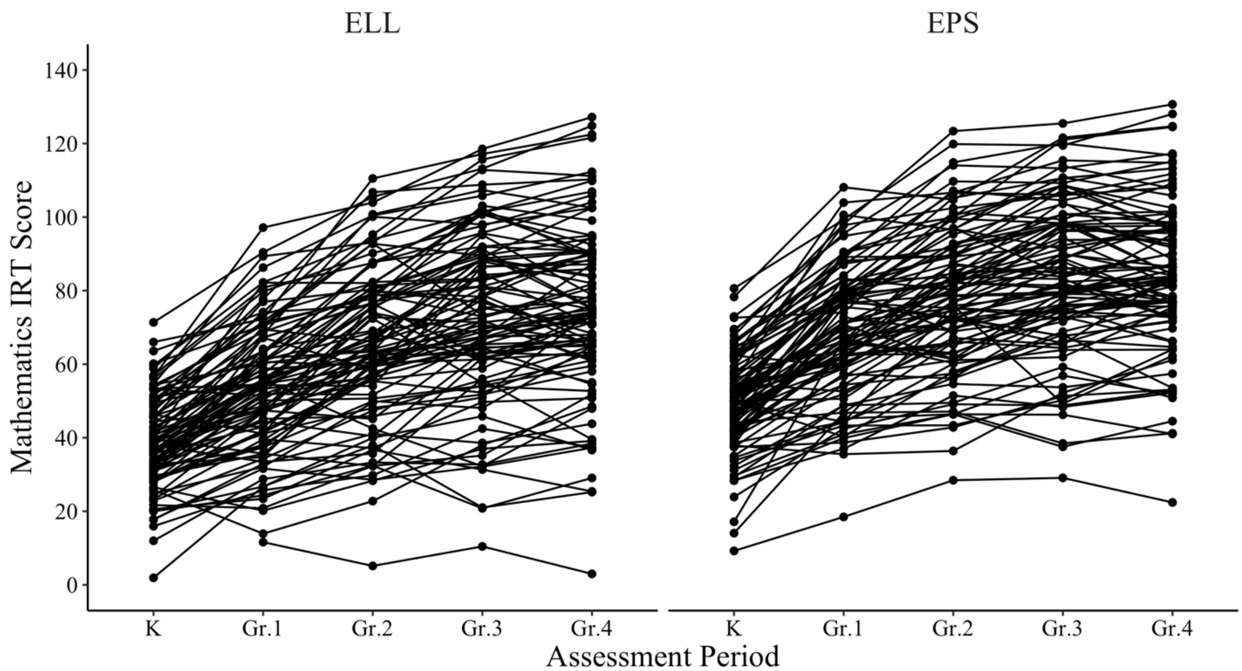
Weighted Boxplots of Fall 2010 Kindergarten Sample preLAS Scores



Note. Parents of students who reported more than one primary home language were included in "Yes" for the purposes of this study.

Figure S3

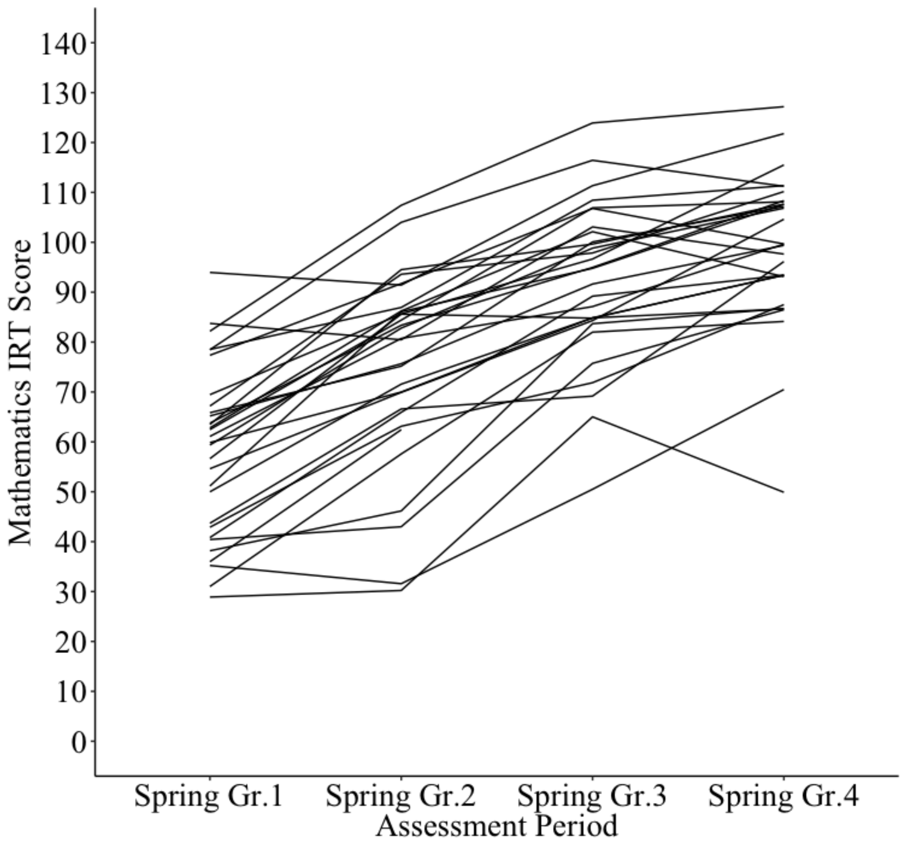
Simulated Trajectories Based on Figure 1 Model Parameters



Note. 100 trajectories simulated for each group. Negative Y values are simulated from the model, though these are not plausible in the actual data and are removed from the graph. Residual variances of the model included in the trajectory simulation. ELL= English langauge learner, EPS = English-proficient student. Plot produced in *ggplot2* (Wickham, 2016).

Figure S3

Random Sample of 30 ELLs' Mathematics Trajectories from Grades 1–4



Note. Plot produced in *ggplot2* (Wickham, 2016).