2023

Stability of Learning Disabilities, Cognitive Growth, and L1 in English Learners: A Latent Class and Transition Analysis

H. Lee Swanson
University of New Mexico

Jennifer E. Kong
Chapman University, jekong@chapman.edu

Stefania D. Petcu
University of New Mexico

Recommended Citation

Swanson, H. L., Kong, J., & Petcu, S. D. (2023). Stability of learning disabilities, cognitive growth, and L1 in English learners: A latent class and transition analysis. *Journal of Educational Psychology, 115*(3), 379–404. https://doi.org/10.1037/edu0000771

Stability of Learning Disabilities, Cognitive Growth, and L1 in English Learners: A Latent Class and Transition Analysis

Comments

This is a pre-copy-editing, author-produced PDF of an article accepted for publication in *Journal of Educational Psychology*, volume 115, issue 3, in 2023 following peer review. The definitive publisher-authenticated version is available online at https://doi.org/10.1037/edu0000771.

Stability of Learning Disabilities, Cognitive Growth, and L1 in English Learners:

A Latent Class and Transition Analysis

H. Lee Swanson

University of New Mexico and University of California-Riverside

Jennifer Kong

Chapman University

Stefania D. Petcu

University of New Mexico

H. Lee Swanson, Ph.D., Educational Psychology, University of California-Riverside (Lee.Swanson@ucr.edu) and University of New Mexico (HLswanson@unm.edu). Jennifer Kong, Ph.D., Special Education, Chapman University and post-doctoral University of New Mexico (JenniferKong@unm.edu) and Stefania D. Petcu, Ph.D., Department of Individual, Family and Community Education, University of New Mexico (spetcu@unm.edu).

This research was supported by an NSF grant, Division of Research on Learning (award number 1660828) Sciences awarded to the first author. This study does not necessarily reflect the views of NSF, or the participating school districts. Special appreciation is given to Monica Ascencio Pimentel, Mario Del Angel Guevara, Lorena Luevano, Julianna Massa, Catherine Riskie, Yuliana Kenfield,Miles Brooks, Hope Gill, Jui-Teng Li, Genesis Arizmendi, Daniel Dickson, Ana Sanchez, Cinthia Lazcano, Christy Yan, Allison Serceki, Jessica Anazagasty, Ancary Ramirez, Brenda Tavares, Oscar Villasenor, Vanessa Valenzuela,Karen Cruz-Zuniga, Selina Garcia and Bernadette Hall-Cuaron and School District Liaisons and Consultant: Rosie Gonzales and Erin Bostick Mason for data collection and/or analysis.
Correspondence concerning this article should be addressed to H. Lee Swanson, Educational Psychology, College of Education, University of New Mexico, and Albuquerque, NM 87131.

All materials for this study can be obtained from the authors or at the e-mail below.

E-mail:HLswanson@unm.edu or Lee.Swanson@ucr.edu

Revision Submitted: 6/21/2022

Abstract

This study investigated the stability of latent classes of students with learning disabilities among a heterogeneous sample of elementary-aged children whose first language is Spanish. To this end, children (*N* = 284) in Grades 1, 2, and 3 at Wave 1 (Year 1) were administered a battery of vocabulary, reading, math, and cognitive measures (short-term memory, working memory, rapid naming, inhibition) in both Spanish (L1) and English (L2). These same measures were administered one and two years later (Wave 2 and 3). Two stable latent classes of children at risk for learning disabilities (children with comorbid difficulties and children with high order difficulties) emerged that were distinct from two latent classes (balanced bilinguals, unbalanced bilinguals) of average achievers who varied in second language acquisition. Further, significant growth parameters that uniquely predicted the log-odds identifying latent classes across all status groups were measures of working memory. Finally, the significant contributions of L2 cognitive measures to latent class status were dependent on L1 cognitive performance. The results suggest that statistically distinct and stable latent classes of children with learning disabilities emerge under the umbrella of English language learners and that growth in the executive processes of working memory and first language cognitive performance play an important role in predictions of latent class status.

Keywords: reading and math disabilities, working memory, learning disabilities, latent transition analysis, English learners

Educational Impact and Implications

English learners with Spanish as a first language bring many assets and strengths to L2 (second language) reading and math tasks, which may transfer from their proficiency in their L1 (first language). However, confounds exist in the assessment of children with potential learning problems who are second language learners who could be provided intervention services. Some of these confounds are due in part to attributing difficulties in second language acquisition and reading/math achievement to the same academic skills and cognitive processes as children with learning disabilities. Other confounds relate to only testing the child in their second language. The present study identified two discrete and stable latent classes of English learning (EL) children at risk for learning disabilities (LD) in both language systems separate from average achieving children who varied in second language acquisition. Key cognitive measures in the predictions of latent class status were naming speed and working memory Specific cognitive inefficiencies in L1 were related to cognitive inefficiencies in L2. The present findings suggest there are several limitations of longitudinal studies that assess risk only in the child's second language.

Stability of Learning Disabilities, Cognitive Growth, and L1 in English Learners: A Latent Class and Transition Analysis

English learners (ELs) are a heterogeneous group of students who speak, read, write or understand a language other than English in their homes. In U.S. public schools, ELs with Spanish as the first language constitute a large percentage of the EL population (73.1%), and are more likely to have reading and/or math difficulties in the elementary school years than other children learning English as a second language (e.g., Batalova & McHugh, 2010; Hemphill & Vanneman, 2011; National Assessment of Educational Progress [NAEP], 2011; 2017, 2020; see Kieffer & Thompson, 2018, for a critical review). Compounding these aforementioned difficulties is that many of these EL children with reading and/or math difficulties are not provided appropriate intervention services (e.g., Artiles et al., 2006; Mancilla-Martinez et al., 2019; Morgan & Farkas, 2016; Sullivan et al., 2015). Clearly, not all children with Spanish as a first language and learning English as a second language will experience learning difficulties and therefore need intervention services. English learners with Spanish as a first language bring many assets and strengths to reading and math tasks, which may transfer from their proficiency in their L1 (first language). However, confounds do exist in the assessment of children with potential learning problems who are second language learners who could be provided intervention services (e.g., Francis et al., 2019; Santi et al., 2019; Rojas et al., 2019). Some of these confounds are due in part to attributing difficulties in second language acquisition and reading/math achievement to the same academic skills and cognitive processes (e.g., Cho et al., 2019; Farnia & Geva, 2019; Goodrich & Namkung, 2019; Swanson, Kong & Petcu, 2019).

The aforementioned issues call for better tools and methods for accurately identifying

EL children at risk for reading and/or math difficulties. There have been some notable methodological advances made to identify discrete learning groups, such as children with learning disabilities (e.g., Grimm et al., 2019) based on latent class analysis (e.g., Lanza et al., 2013; Muthén & Asparouhov, 2020). Latent class analysis (LCA) is a statistical method used to identify subgroups of individuals characterized by similar multidimensional patterns of responses (e.g., Collins & Lanza, 2010). Latent class modeling can be used to identify group membership of specific classes of individuals based on proficiency in reading and/or math. Although reading and/or math performance in EL children can be represented as a continuous outcome variable, the sample may be composed of different groups (or classes) of individuals. This group membership is not directly observed in latent growth models even though the distribution of children's reading and/or math proficiency may reflect at least two different latent classes (e.g., children at risk and not at risk for learning disabilities). Unfortunately, when focusing only on continuous reading and/or math variables, the distributions of these latent classes are not observed, even though the continuous variable may reflect a mixture of distributions (see Hickendorff et al., 2018, Nylund-Gibson & Choi, 2018; Vermunt, 2007; for review).

Several studies have used latent growth models to better understand children at risk for learning disabilities (e.g. Peng et al., 2016b; Quinn et al., 2020). Although latent growth modeling permits the researcher to use the rate at which the processes change over time, it does not tell us much about the unidimensionality or the state of EL children at risk for RD and/or MD. That is, the state of risk for learning disabilities may be fairly stable or extremely transient. As an example, students may not be identified as at risk for learning disabilities in grade 1, but experience difficulties in grade 3. Latent transition analysis (LTA) is well suited for the study of latent variables that have the potential of changing over time, as it can be used to test models involving a stage sequential development; this procedure differs from LCA in which individuals do not change membership over time.

To the best of our knowledge, only a few studies have addressed these aforementioned issues related to EL children using latent class and latent transition analysis on an array of language, achievement, and cognitive measures (Catts et al., 2012; Grimm et al., 2019; Swanson et al., 2016; 2021). An earlier study (Author et al., 2016) identified a distinct latent class of EL children at risk for reading difficulties at the first testing wave that were also exhibiting difficulties three years later. The results indicated an extremely low incidence of children identified as good readers in wave 1, being identified as at risk for reading difficulties in the later testing waves. There were several limitations in this study, however, which a follow-up study addressed (Author et al., 2021). The primary limitation of the earlier work was focusing on EL children at risk for reading difficulties defined on English measures of reading. That is, if risk status for some EL children is related to learning disabilities, which in turn is assumed to have a biological base (see footnote 1), then it must be shown that the disabilities in reading/or math do not reside in just one language system. Such was not the case in this study.

A more recent study (Swanson et al., 2021) considered the classification of elementary children at risk for learning disabilities based on performance on English and Spanish normed referenced measures of vocabulary, reading, and mathematics. In addition, the transitional probabilities between grades at two test waves were analyzed. Discrete groups of EL children at risk for both reading disabilities and math disabilities (MD) emerged. However, the stability of these latent classes could not be assessed. Although some classes were stables across the two time periods others were not.

These earlier studies, however, highlighted some of the difficulties in identifying EL children at risk for learning disabilities and whether those disabilities are stable over time. One of the difficulties was the reliance on only two-time points (e.g., Swanson et al., 2021). One cannot assume that stability over time is isolated to two measurement occasions (Hickendorff et al., 2018, p. 12). Further, the shape of data (e.g., linear, quadratic, curvilinear) could not be determined (e.g., Rogosa & Gottman, 1995). Perhaps, a more serious difficulty is that children exhibiting academic difficulties in one grade will not exhibit academic difficulties in the next grade and vice versa. For example in the domain of math, Morgan et al. (2009) examined the stability of MD as a function of low performance at various time points. The stability of MD from kindergarten to Grade 5 ranged from 28% to 65% of the sample. Likewise, Jordan and Hanich (2003) reported that only 18% of students identified as experiencing MD in Grade 2 retained the same status in Grade 3; however, they also reported that 88% of students identified as experiencing comorbid math and reading difficulty retained the same classification. Other issues related to determining a stable latent class of EL children at risk for learning disabilities reflect changes in definitional criteria (e.g., cut-off scores related to defining risk status change), curriculum changes, and/or demands placed on various language, achievement, and/or related cognitive processes in the subsequent grades (e.g., see Etmanskie et al., 2016; Martin et al., 2013, for review).

To address some of these issues, the current study has three purposes. The first purpose determines if children at risk for learning disabilities within a heterogeneous EL sample reflect a stable latent class across three testing waves. The second purpose determines if distinct cognitive processes predict children at risk for learning disabilities from children who vary in language acquisition. The final purpose determines whether cognitive processing in L1 has an impact on

L2 processing in determining latent class. The literature addressing each of these purposes follows.

Language Acquisition vs. Learning Disability

Children have been defined as at risk for specific learning disabilities by performing below a cut-off score point on norm-referenced standardized reading and/or math measures (e.g., Brandenburg et al, 2017; Branum-Martin et al., 2013). For example, performance below the 25th percentile on standardized norm-referenced reading and/or math measures across several grades is a common standard to identify children at risk for reading difficulties (RD) and/or math difficulties (MD) (e.g., Fuchs et al., 2006; Geary et al., 2012; Murphy et al., 2007; Siegel & Ryan, 1989; Stanovich & Siegel, 1994; Yeung, 2018). Multiple measures of reading and math are included in the identification process because children with reading and math difficulties share similar cognitive processing difficulties and/or common dimensions that underlie their risk status (Child et al., 2019; Mann, & Miller, 2013, Swanson, 2020; Swanson & Fung, 2016). [1]

No doubt, several criticisms emerge as to identifying children at risk of learning disabilities because of a reliance on arbitrary cut-off scores (e.g., Branum-Martin et al., 2013; Cirino et al., 2015). Although latent class analysis or mixed analysis allows for the determination of discrete latent classes within these cut-off points, these cut-off points have also been exacerbated when defining risk status among EL students because such children are not tested in their first language (e.g., Peña et al., 2016). This is because children who are learning two languages must attain a threshold in their first language before cognitive and academic skills can be assessed, as a child's mastery of their second language is a function of the child's linguistic competence in their first language (e.g., Cummins), 1979; however see Prevoo et al., 2016, for a meta-analysis of this literature).

In response to these issues, the current study includes a broad range of manifest variables administered over three years to identify difficulties in reading and math on both basic [word identification, calculation] and high order tasks [reading comprehension and mathematical problem solving] within both language systems. The study determines if learning disabilities appear as a stable latent class or merely a function of variation in second language acquisition.

Underlying Cognitive Processes

Current procedures to identify children with potential learning disabilities assume that such children experience cognitive constraints, which impede their ability to perform efficiently on language and achievement measures (e.g., Geary et al, 2017; Kaushanskaya, & Yoo, 2013; Lesaux et al., 2006). Working memory has been identified as one of the most often referred to cognitive processes underlying both RD and MD (WM; David, 2012; Peng et al., 2016a, 2016b, 2018; Swanson et al., 2004), which has also been related to achievement difficulties in EL children (e.g., Swanson et al., 2006; 2015). Working memory (WM) is defined as a limited capacity system of temporary stores, functions related to the preservation of information while simultaneously processing other information, and attention control related to these functions (e.g., Baddeley, 2012).[2] Cowan (2014) defines WM "as the small amount of information that can be held in the mind and used in the execution of cognitive tasks, in contrast with long-term memory, which is the vast amount of information saved in one's life "(p.197). What is not apparent from the research is whether a general executive (WM) system or a specialized system (STM) captures academic performance across a large range of academic difficulties in EL children.

Working memory has been shown to contribute significant variance to EL children's reading and math achievement, even when domain-specific processes related to reading (e.g., phonological awareness; Swanson et al., 2015) and math (e.g., estimation, numeracy; Swanson et al. 2019) are included in the regression modeling. Although the association between WM and reading and/or math has been established in the literature, the processes of WM that underlie predictions of reading and/or math performance are unclear (see Peng et al., 2016a; 2016b; 2018; for review). Some studies have suggested that the storage component of WM (referred to as verbal short-term memory, STM, or the phonological loop) plays a major role in academic performance (e.g., Peng et al., 2016b). Other studies have noted that academic difficulties are tied to the executive component of WM (e.g., Swanson et al., 2015). Although WM or complex span tasks share the same processes (e.g., rehearsal, updating, controlled search) as short-term memory (STM) or simple span tasks, simple tasks have a greater reliance on phonological processes than WM or complex span tasks (see Unsworth & Engle, 2007, pp. 1045-1046, for a review). Thus, we considered two cognitive models in accounting for language and academic proficiency in EL children at risk for learning disabilities: one focuses on processing efficiency at a phonological level and the other focuses on executive processes.

Phonological system. Several studies support the notion that phonological processes underlie basic skills in reading and math computation (e.g., McBride-Chang et al., 2012; Verhagen & Leseman, 2016). Indeed, one of the components of WM commonly attributed to academic and language development is the phonological loop (phonological STM; Baddeley et al., 1998; de Abreu & Gathercole, 2012). Because the phonological loop has been found to play a major role in second language acquisition and academic achievement of EL children (e.g., de Abreu & Gathercole, 2012; Foster et al.,2015; González & Valle, 2000; Gorman, 2012), it could be argued that difficulties in higher-order processes, such as reading comprehension or math problem solving, are an artifact of phonological storage. That is, the phonological system, via the phonological loop (i.e., phonological store; subvocal rehearsal), influences children's verbatim memory capacity, which in turn supports reading and math performance. These assumptions are consistent with some bottom-up models of higher-order processing which view the primary task of executive processing as one of relaying the results of lower-level linguistic analyses upward through the language system. Phonologically analyzed information is transferred to WM storage, which is then transferred (thus freeing storage for the next chunk of phonological information) upward through the processing system to promote the online extraction of meaning. One of the key mechanisms related to the influence of phonological STM on reading and math tasks is naming speed (Georgiou et al., 2013; McBride-Chang et al., 2012). Naming speed has been considered a measure of how quickly items can be encoded and rehearsed within the STM system (e.g., Bonifacci et al., 2011; Georgiou et al., 2013).

Executive system. An alternative model suggests that cognitive processes separate from the phonological system, especially those related to the central executive controlling system of WM, play an equally important role in the prediction of potential learning disabilities. Thus, in contrast to the aforementioned hypothesis that suggests the phonological system plays the dominant role in EL children's academic performance, a second model assumes that executive processes (i.e., controlled attention) also play a major role. The influence of WM on measures of reading and math would follow automatically with improvements in controlled attention. Controlled attention is operationally defined as the residual variance of a complex WM task that predicts the criterion task (e.g., reading) when phonological storage (STM) has been partialed out in the analysis (e.g., Bayliss et al., 2003; Engle et al., 1999). Two findings underlie this assumption: (1) the executive component of WM has been associated with both reading and math difficulties (e.g., David, 2012; Martin et al., 2013; Menon, 2016) and (2) proficiency in L1 and

L2 positively influence controlled attention, cognitive flexibility, and overall executive processing (e.g., Bialystok, 2011).

Thus, the influence of the executive component of WM is related to one's ability to accurately access information from LTM, monitor possible action sequences (e.g., associative links), and update information (e.g., Engle, 2018). A related mechanism in this process is the inhibition of competing language systems (Bialystok, 2011; Bialystok & Feng, 2009; Bonifacci et al., 2011). Because inhibition has been attributed to WM (e.g., Friedman et al., 2007; Getzmann et al., 2018), as well as reading and math difficulties (e.g., Passolunghi & Siegel, 2004; Toll et al., 2011), individual differences related to WM may play an important role in EL children at risk for RD and/or MD.

Contribution of Spanish (L1) to English (L2) Cognitive Processing

Does L1 (Spanish in this case) have an impact on L2 (English in this case) processing in determining a latent class for EL children at risk for learning disabilities? This is an important question because a large majority of longitudinal studies designed to determine risk test EL children in their second language without the inclusion of L1 measures in the analysis (e.g., Finders et al., 2021; Morgan et al 2017; Vukovic & Lesaux, 2013). The distinction between language acquisition and learning difficulties cannot be assessed unless the role of the first language in achievement outcomes is addressed. Thus, the question emerges as to whether L2 academic difficulties are heightened by difficulties in their first language or can cognitive difficulties that potentially underlie the child's second language operate independently of their first language? This is an important theoretical question as to whether access to both language systems in EL children provides some advantages in academic performance. This is also an important practical question, considering that it has been argued that bilingual instruction has an additive effect on academic performance when compared to only providing training in reading and/or math in one language system (e.g., Lindholm-Leary, 2005; Lindholm-Leary & Block, 2010). The benefits of monitoring two languages have been noted in bilingual children when compared to monolingual children on cognitive measures, such as executive processes (e.g., Bialystok & Feng, 2009; also see Gunnerud et al., 2020). However, the research is unclear as to whether switching or transferring one form of knowledge to another plays a role in learning difficulties in the child's second language (Choi et al., 2018; Kempert et al., 2011).

The interchange between the two language systems among EL children in the present study was assessed by comparing L2 cognitive predictors (e.g., STM, naming speed) of latent class status with and without L1 predictors entered into the regression model. The relative contribution of performance in the Spanish language is tested by comparing the standardized beta weights of the predictors in the equation before and after the inclusion of these measures. Although some L1 measures may not contribute significant variance to overall predictions of latent class status, these L1 processes may play a major role in enhancing the predictions of L2 measures. Thus, in this context, Spanish measures may serve as a suppressor variable (cf. Conger, 1974) or what has also been referred to in the literature as enhancer variables (e.g., McFatter, 1974). Several studies have demonstrated the value of well-conducted suppressor analysis (e.g., Gaylord et al., 2010; Kim, 2019), and we extend this research by looking at how cognitive variables related to the first language play a key role in second language predictions of latent class status.[3]

Taken together, three questions directed this study.

1. Can a stable latent class of EL children be identified with learning disabilities among children who vary in language acquisition?

2. Do measures of growth and the level of cognitive performance predict the latent class status?
3. Do Spanish cognitive processes enhance or operate independently of English cognitive predictions of latent class status?

In summary, the present study determines whether a stable latent class emerges related to risk for learning disabilities in reading and/or math risk among EL children. Also of interest are the potential cognitive measures that predict the latent classes as well as the role cognitive processing in the child's first language (Spanish) plays in such predictions. Thus, to identify latent classes at risk for learning disabilities, a battery of Spanish and English normative measures in vocabulary, reading and math was administered across three years. We assumed that EL children at risk for learning disabilities experience academic difficulties in **both** language systems across the three testing waves. In contrast, children with average vocabulary and academic achievement within one language system but not the other were assumed to experience potential difficulties related to second language acquisition. To determine children at risk for learning disabilities it was necessary to establish that (a) such children's academic difficulties were not due to (a) general intellectual difficulties (e.g., Ferrer et al., 2010) and (b) comorbid difficulties, such as those with Attention-Deficit/Hyperactivity Disorder (ADHD; e.g., Boada et al., 2012; Snowling, 2012). Thus, measures of fluid intelligence and behavioral ratings of attention were also included in the classification battery. Measures of attention were included because children with ADHD are viewed as having primary processing difficulties that are distinct from children with learning disabilities (e.g., Willcutt et al., 2005). Further, children with learning disabilities are assumed to be in the normal range of intelligence and this represents a distinct category of children from those with general intellectual challenges (e.g., Giofrè et al., 2017).

Method

Participants

This study was approved by the Institutional Review Board at Univeristy of New Mexico, Project title: Math problem solving and working memory growth I English learers at risk for Math Disabilities, IRB protocol:99581-1. The study included a sequential-cohort design in which children in Grades 1,2,3 were assessed in Years 1, 2, and 3, creating three cohorts (Cohort 1: Grades 1-3, Cohort 2: Grades 2-4, Cohort 3: Grades 3-5). Two hundred eighty-four (N=284) students in grades 1 (n =115), 2 (n =91), and 3 (n =78) from two large urban school districts in the southwest United States participated in this study. The sample was part of a larger federally funded longitudinal study assessing cognitive growth among EL children for a three-year period (Swanson et al., 2021). Children who were tested in reading and math for at least two testing waves in both English and Spanish were the focus of this analysis.

The children in this study were designated by their school as English learners (ELs) and were selected from 31 elementary classrooms at Wave 1. EL status in the schools was determined by a language usage survey in the home and a state-approved English language proficiency (ELP) screener. School records indicated children's primary home language was Spanish (> 95%). The children were later tested and were nested in 23 classrooms at Wave 2 and 23 classrooms at Wave 3. The complete sample included 132 boys and 152 girls who returned signed consent forms. No significant differences in gender representation emerged at wave 1, χ^2(df=2, N= 284) = .76, p = .68, or wave 3, χ^2(df=2, N= 266) =1.32, p = .52. Thus, there was some attrition at wave 3 (N=18) due to children leaving the school district and/or COVID restrictions. The largest reduction of participants in wave 3 occurred in the middle cohort of children (Cohort 1 N=6, Cohort2 N=10, Cohort 3= N=2). Estimation procedures to address "missingness" are discussed below.

Children in the sample participated in a full or reduced Federal lunch program and were drawn from neighborhoods with high Hispanic/Latino representation. The sample was drawn from four large elementary urban schools in two large metropolitan areas. Two of the elementary public schools in the sample yielded, according to state reports, the lowest percentage of reading and math score proficiency within the state. Minority (Hispanic/Latino) enrollment was 95% of the student body that was higher than the state average. The current study also included two urban charter schools with a high Hispanic/Latino (> 95%) representation. State reports indicated that one of the charter schools at the time of testing (2017-2018) reported that only 35% of children were proficient in reading and 29% were proficient in math. A state report on the second charter school had also indicated that only 33% of the elementary children were proficient in reading and 29% proficient in math on state measures. All students attended a Dual Language Immersion program (two-way immersion), following the 90/10 Model. In this model, instruction is delivered in 90% Spanish and 10% English in Kindergarten in all subject areas, increasing annually until both English and Spanish are taught equally. For our sample, specifically, instruction was delivered in 50% English and 50% Spanish by the fourth grade. Although students may have received supplemental instruction in either reading and/or math, intervention for such children was delivered by the school was independent of our data collection. However, six students in our sample had individualized educational programs (IEPs). According to parent reports, approximately 53% of the sample spoke only Spanish, 13% spoke both English and Spanish, and 33% spoke only English at home.

Measures Used for Identifying Latent Classes

The study included group and individual administrations of a battery of tests. The series of tests were counterbalanced into one of four presentation orders. No Spanish and English versions of the same test (except for the Expressive One-word Picture Vocabulary Test, Spanish-Bilingual Edition; Brownell, 2001) were presented simultaneously. All participants were administered both English and Spanish versions of each measure by bilingual graduate students and staff researchers. Because the norm-referenced measures for establishing the latent classes are commercially available, along with information on their validity and reliability, the measures are only briefly reviewed here. Additional detail was provided below for the experimental cognitive measures.

Vocabulary: Receptive and Expressive

Peabody Picture Vocabulary Test (PPVT*).* The Peabody Picture Vocabulary Test (Dunn & Dunn, 2007) was administered in English. In this task, children were presented with four pictures and asked to select the picture that matched the word read aloud in English. Word presentation gradually increased in difficulty.

Test de Vocabulario en Imágenes Peabody (TVIP*).* This measure is similar to the PPVT in the presentation and administration, except that words are read aloud in Spanish (Dunn et al., 1986). The sample reliabilities (KR_{20}) in English (PPVT) were .97, .96, .96 and Spanish (TVIP) was .92, .94, and .92 for waves 1, 2 and 3, respectively.

Expressive One-Word Picture Vocabulary Test - Spanish-Bilingual Edition. The Expressive One-Word Picture Vocabulary Test - Spanish-Bilingual Edition (EOWPVT-SBE; Brownell, 2001) was used as a measure of English and Spanish expressive vocabulary, in which children were presented with a picture and were asked to name the object in the testing language.

The sample reliabilities (KR_{20}) in English were .96, .95, .95 and Spanish were .95, .95, and .93 for waves 1, 2 and 3, respectively.

Math: Calculation and Word Problem-Solving.

Arithmetic calculation. The calculation subtests from the *Woodcock-Johnson III* (Woodcock et al., 2001) were administered for the English presentation, and Cálculo from the *Batería III Woodcock-Muñoz* (Muñoz-Sandoval et al., 2005) was administered to establish norm-referenced math levels in Spanish. The subtest required written computation of problems that increased in difficulty. The sample reliabilities (KR_{20}) for WJ calculation measures in English were .76, .80, .82 and Spanish were .70, .68, and .68 for waves 1, 2 and 3, respectively.

Math word problem-solving. The math applied problem-solving subtest from the *Woodcock-Johnson III* (Woodcock et al., 2001) was administered for the English presentation and the *Problemas Aplicados* from the *Batería III Woodcock-Muñoz* (Muñoz-Sandoval et al., 2005) was administered to establish norm-referenced math levels in Spanish. Both of these subtests were individually administered and assessed children's early mathematical operations (e.g., counting, addition, and subtraction) through practical problems. To solve each math word problem, the subject listened to the formulation, recognized the procedures that must be followed, and then performed relatively simple calculations. The sample reliabilities (KR_{20}) in English were .76, .84, .89 and Spanish were .83, .67, and .85 for waves 1, 2 and 3, respectively.

Reading: Word Identification and Passage Comprehension.

Woodcock-Muñoz Language Survey-Revised (WMLS-R). The WMLS-R Spanish and English word identification and passage comprehension subtests were administered. This test assessed children's reading levels in English and Spanish (Woodcock et al., 2005). Depending on the language of administration, the subtests were administered individually to students in English or Spanish.

Word Identification. For the word identification subtest, children were tested individually by presenting them with a list of words, which gradually increased in difficulty. The words followed regular spelling patterns in this non-timed test. The sample reliabilities word identification (KR_{20}) in English was .94, .92, .92, and Spanish were .90.92, .93 for waves 1, 2, and 3, respectively. The sample reliabilities word identification (KR_{20}) in English was .94, .92, .92, and Spanish were .90.92, .93 for waves 1, 2, and 3, respectively.

Passage Comprehension. For the passage comprehension subtest, children identified specific words that go in the blank spaces of various passages. Early passages were accompanied by a corresponding picture and sentences gradually increased in complexity. The sample reliabilities for passage comprehension (KR_{20}) in English were .83, .86, .80 and Spanish were .84, .81, and .81 for waves 1, 2 and 3, respectively.

Fluid intelligence and attention.

Fluid (nonverbal) intelligence. The Colored Progressive Matrices test (Raven, 1976) was used as an indicator of nonverbal reasoning or fluid intelligence. Children were given a booklet with patterns on each page, each pattern revealed a missing piece. For each pattern, six possible replacement pattern pieces were displayed. Children had to select (circle) the correct missing piece of the pattern. Instructions were presented in both English and Spanish. The sample reliabilities (KR_{20}) were .81, .83, and .86 for waves 1, 2 and 3, respectively.

Classroom Attention. Attention was assessed by administering the Conners' Teacher Ratings Scales-Revised: Short Form (CTRS–R:S; Conners, 1997). The CTRS-R:S is used to evaluate students' problem behaviors by obtaining ratings from teachers. The homeroom teacher was selected for each child and was asked to complete the CTRS–R: S. The primary measure for this study was the Attention-Deficit/Hyperactivity index.

Cognitive Measures Used for Determining Correlates of Latent Class Membership

The cognitive measures assumed to be related to the latent classifications assessed the storage of phonological information (short-term memory and naming speed) and executive processing (central executive of working memory, visual-spatial working memory, and inhibition). The convergence of the measures for the English and Spanish versions was established in an earlier study (see Swanson, Kudo, & Van Horn., 2019, for further discussion), and a full description of each cognitive measure is provided in Swanson et al. (2004; 2008; 2015, 2019).

The cognitive measures included a translation from English to Spanish as well as a backward translation process (re-translating content from the target language back to its source language in literal terms). Although raw score measures were converted to z-scores and latent measures (see below) so that responses were on the same scale (M=0, *SD*=1), English and Spanish words and numbers were not balanced (or could not be balanced) for length (e.g., syllables), because Spanish task words and numbers are longer on average. The translation could not control for the number of syllables, frequency of use, imagery, and meaning across measures. Spanish and English instructions, however, included practice items, and the test items corresponded with expected language responses during individual testing. Thus, the backward translation of the items may have led to more memory demands placed on Spanish tasks. However, as shown in supplemental Table 3, for the total sample, memory scores for both English and Spanish yielded moderate correlations suggesting some shared variance in terms of difficulty. In addition, a comparison of effect sizes between average achievers relative to the other groups showed larger effects sizes on English than Spanish memory measures. However, it is important to note that the difficulty of processing items in one language versus another could not be controlled and may account for some of the outcomes.

Phonological Storage and Naming Speed

Short-term memory (STM) measures (phonological loop). Four tasks of STM were administered in Spanish and English: forward digit span, backward digit span, word span, and pseudoword span. The forward and backward digit span tasks (taken from the WISC-III; Wechsler, 1991) and Spanish-translated versions of these two tasks were administered. The forward digit span task required children to recall sequentially ordered sets of digits that increased in number. The backward digit span task required children to recall sets of digits, but in reverse order. Backward digit span was included as part of STM since it loads on the same factor as forward digit span and is associated with phonological memory (e.g., Colom et al., 2006). Dependent measures for both tasks were the largest set of items recalled in order (range = 0 to 8 for forward digit span; range = 0 to 7 for backward digit span). For the translated Spanish versions of the forward and backward digit span subtests, identical numbers were presented in the same order as the English version. There were no deviations in the procedure, except for the language used. The sample reliabilities (KR_{20}) for forward digit in English were .41, .58, .78 and Spanish were .74, .65, and .55 for waves 1, 2 and 3, respectively. The sample reliabilities (KR_{20}) for backward digit in English were .79, .51, .61 and Spanish were .66, .49, and .45 for waves 1, 2 and 3, respectively.

The word span and pseudoword span tasks were presented in the same manner as the forward digit span task. During the word span task, examiners read lists of one or two-syllable, high-frequency words that included unrelated nouns and then asked the children to recall the words. The pseudoword span task (phonetic memory span task) used strings of one-syllable nonsense words, which were presented one at a time in sets of two to six nonwords (e.g., DES, SEEG, SEG, GEEZ, DEEZ, DEZ). A parallel version was developed in Spanish for the word span and pseudoword span tasks. The dependent measure for all STM measures was the highest set of items retrieved in the correct serial order (range 0 to 7). The sample reliabilities (KR_{20}) for word span in English were .76, .74, .67, and Spanish were .76, .74, and .73 for waves 1, 2 and 3, respectively. The sample reliabilities (KR_{20}) for pseudoword span in English were .77, .67, .69 and in Spanish were .74, .68, and .92 for waves 1, 2 and 3, respectively.

Rapid naming of digits and letters. Several models assume that operations related to academic achievement (e.g., math) are time-related (e.g., Bonifacci et al., 2011; Georgiou et al., 2013), and therefore naming speed was assessed in this study. The administration followed those specified in the manual of the Comprehensive Test of Phonological Processing (CTOPP; Wagner et al., 2000). For this task, the examiner presented the child with an array of items (e.g., six letters or six digits). Depending on the day of administration, children were asked to name the items, speaking in English or Spanish, as quickly as possible for each stimulus set. The dependent measure was the amount of combined time it took for children to complete each set. The number of errors was also taken into account in creating the final score. The sample reliabilities (KR_{20}) for letters in English were .94, .91, .89 and Spanish were .95, .91, and .95 for waves 1, 2 and 3, respectively. The sample reliabilities (KR_{20}) for numbers in English were .94, .90, .85 and Spanish were .93, .89, and .89 for waves 1, 2 and 3, respectively.

Inhibition and Executive Processing.

Inhibition. The use of random generation tasks has been well articulated in the literature as a measure of inhibition (e.g., Towse & Cheshire, 2007). The task required children to actively monitor candidate responses and suppress responses that would lead to well-learned sequences, such as 1-2-3-4 or a-b-c-d (Baddeley, 2007). For this study, each child was asked to write numbers (or letters) as quickly as possible, first in sequential order and then in random non-systematic order. For example, children were first asked to write numbers from zero to ten in order (i.e., 1, 2, 3, 4) as quickly as possible within a 30-second period. They were then asked to write numbers as quickly as possible "out of order" within a 30-second period. After correcting for information redundancy and the percentage of paired responses, the dependent measure was the number of random items written. Children performed these tasks in English or Spanish, on separate days. The sample reliabilities (KR_{20}) for letters in English were .70, .70, .74 and Spanish were .78, .78, and .71 for waves 1, 2 and 3, respectively. The sample reliabilities (KR_{20}) for numbers in English were .58, .68, .72 and Spanish were .66, .74, and .79 for waves 1, 2 and 3, respectively.

Central executive of WM. The conceptual span, listening sentence span, digit-sentence, and updating task were administered in English and Spanish to capture the executive component of WM. These tasks required children to hold increasingly complex information in memory while simultaneously responding to a question about the task. For example, after children listened to a list of words in the conceptual span task, they were asked, "Which word from the list did I say, X or Y?" They were then asked to recall words from the list. This balance of simultaneous storage and processing is consistent with several studies of WM processing, including Daneman and Carpenter's (1980) seminal WM measure.

The conceptual span task was used as an indicator of WM processing that involves the ability to organize sequences of words into abstract categories. Children listened to a set of words that, when re-organized, could be grouped into meaningful categories. For example, they were told a word set such as "shirt, saw, pants, hammer, shoes, nails." After answering a distracter question, they were asked to recall the words that "go together" (i.e., shirt, pants, and shoes; saw, hammer, and nails). The range of set difficulty was two categories containing two words each, up to four categories with four words each. A Spanish- translated version was also administered on a separate day. Care was taken in the development of the measure to keep the abstract categories the same in both languages (e.g., clothes and tools). The dependent measure for both versions was the number of items recalled correctly. The sample reliabilities (KR_{20}) for conceptual span in English were .74, .69, .70 and Spanish were .73, .60, and .74 for waves 1, 2 and 3, respectively.

The children's adaptation (Swanson, 2013) of Daneman and Carpenter's (1980) listening sentence span task was also administered. This task required the presentation of groups of sentences, read aloud, for which children tried to simultaneously understand the sentence contents and remember the last word of each sentence. The digit sentence task measured the participant’s ability to recall numerical information that was embedded within a short sentence. The numerical information referenced either a location or address. On this test, the examiner read a sentence and then asked the examinee a process question. The examiner then asked the examinee to recall the numbers in the sentence. Children performed these tasks in English and Spanish, on separate days. The sample reliabilities (KR_{20}) for listening span in English were .51, .59, .77 and Spanish were .31, .36, and .49 for waves 1, 2 and 3, respectively. The sample reliabilities (KR_{20}) for digit-sentence span in English were .77, .81, .83 and Spanish were .74, .79, and .80 for waves 1, 2 and 3, respectively.

Because WM tasks were assumed to tap a measure of controlled attention referred to as updating (e.g., Miyake et al., 2000), an experimental Updating task, adapted from Swanson et al., (2004), was also administered. A series of one-digit numbers was presented that varied in set length from 3, 5, 7, and 9. No digit appeared twice in the same set. The examiner told the child that the length of each list of numbers might be 3, 5, 7, or 9 digits long. Children were then told that they should only recall the last three numbers presented. Each digit was presented at approximately one-second intervals. After the last digit was presented, the child was asked to name the last three digits, in order. The dependent measure was the total number of sets correctly repeated (range 0 to 16). For the translated Spanish version, identical numbers were presented in the same order as the English version. There were no deviations in the procedure, except for the language used. Children performed these tasks in English and Spanish, on separate days. The sample reliabilities (KR_{20}) for updating in English were .71, .85, .91 and Spanish were .67, .84, and .92 for waves 1, 2 and 3, respectively.

Visual-spatial WM. Some studies have noted that difficulties in solving math word problems are related to measures that tap the visual-spatial sketchpad (e.g., Ashkenazi et al. 2013). Thus, two measures were administered to assess visual-spatial WM: Visual Matrix and Mapping & Directions (Swanson, 1992; 2013). Instructions for these measures were given in both English and Spanish for a single administration of each task. The visual matrix task assessed the ability of participants to remember visual sequences within a matrix. Participants were presented a series of dots in a matrix and were allowed five seconds to study the matrix. The matrix was then removed and participants were asked, in both English and Spanish, "Are there any dots in the first column?" After answering the process question, students were asked to draw the dots they remembered seeing in the corresponding boxes of their blank matrix response booklet. The task difficulty ranged from a matrix of four squares and two dots to a matrix of 45 squares and 12 dots. The dependent measure was the number of matrices recalled correctly (range of 0 to 11).

The mapping and directions task required the child to remember a sequence of directions on a map (Swanson, 1992; 2013). The experimenter presented a street map with dots connected by lines; the arrows illustrated the direction a bicycle would go to follow this route through the city. The dots represented stoplights, while lines and arrows mapped the route through the city. The child was allowed 10 seconds to study the map. After the map was removed, the child was asked a process question [i.e., "Were there any stop lights on the first street (column)?"]. The child was then presented a blank matrix on which to draw the street directions (lines and arrows) and stop lights (dots). Difficulty ranged on this subtest from four dots to 19 dots. The dependent measure was the highest set of correctly drawn maps (range = 0 to 9). The sample reliabilities (KR_{20}) for the matrix task were .95, .95, and .95 and for the mapping/direction task were .69, .70, and .79 for waves 1, 2 and 3, respectively.

Cut-off Point

A plethora of studies that have examined children at risk have used a cut-off score of the 25th percentile (90 standard score) on academic measures for both English monolingual (e.g., Catts et al., 2012; Gottardo et al., 2008; Lipka et al, 2006; Stanovich & Siegel, 1994) and EL children (e.g., Chiappe & Siegel, 2006; Chiappe et al. 2002; Farnia & Geva, 2017; Lesaux et al., 2006). However, as noted in our earlier work (Swanson et al., 2016; 2018), the identification of latent classes below this specific cut-off point does not validate that cut-off point, but rather the results suggest we can identify subgroups to the cut-off to which it was applied. In terms of common cut-off score designations for risk in reading and/or math, the 25th percentile cut-off is frequently used to designate risk, and it is useful to use cut-off scores as practiced in the schools. For example, Fuchs et al. (2008) provided a justification for dichotomization (cut-offs) in the area of learning disabilities (math in this case, see p. 37) to determine risk class by "using a common cut-off score designation".

In sum, the manifest variables (vocabulary, reading, math, fluid intelligence, and attention) used to determine discrete groups were dummy coded as reflecting a normative score below the 25th percentile or at or above the 25^{th} percentile (1 = below the 25th percentile, 2 = at or above the 25th percentile). The 25th percentile (a 90 standard score) was based on the normative data in the test manual from the standardized vocabulary, math, reading, and fluid intelligence measures. The Conners scale was in normative T-scores, with high scores representing higher levels of inattention. Therefore, the 25^{th} percentile on the Conners scale was a T-score of 61.

Procedures

Children were tested individually and in groups after informed consent forms were obtained for participation. Ten bilingual graduate students or research assistants trained in test administration tested all participants in their schools. Each child was tested individually and in small groups of 10 to 15 students after informed consent was obtained for participation. For each testing wave, children participated in two sessions of individual testing that lasted 30 to 60 minutes and two group-testing sessions that lasted approximately 60 minutes and occurred for two consecutive days. Group testing occurred for two consecutive days for approximately one hour each day. Measures that allowed for group administration were calculation, visual matrix, number generation, and Raven Colored Progressive Matrices Test. Individual administration included measures of problem-solving, vocabulary, reading, STM, WM, naming speed, and mapping/directions. One of six presentation orders, related to the individually administered tasks, was randomly assigned to each child to control for order effects. Also, the presentation orders of Spanish and English tests were counterbalanced across all children. For the group-administered tests, the presentation order of English and Spanish measures for each type of task were also counterbalanced across small groups.

Statistical Analysis

Missingness

As shown in Supplement Table 1, several of the tasks had missing data. Some children dropped out of the study at the final testing wave. In addition, some children could not do the tasks (understand English and/or Spanish instructions) in the first testing wave, and therefore instead of assigning a value (e.g., 0 or a low standard score of 55), children's performance was assigned as missing data for that task. For the latent class analysis and latent transition analysis, parameters were estimated by maximum likelihood using the EM algorithm, with Newton-Raphson incorporated into the estimation of regression coefficients (e.g. Lanza et al., 2011). The null hypothesis that data are missing completely at random (MCAR) was tested within each model and is reported in Table 2. A test of the MCAR was determined by dividing the reported log-likelihood for the model by the degrees of freedom. If the probability is not significant (p=~1), the results suggested that data were missing at random (Lanza et al., 2011).

As will be discussed below, the latent transition analysis was followed by a hierarchical logistic regression analysis that included the cognitive variables as predictors of latent class status. Given the non-normal nature of the outcome variable (latent classes), the use of ML estimation was not appropriate. Instead, the Laplace estimation was used (see. See Raudenbush et al., 2000 for discussion). The LaPlace approximation method has performed well in simulation studies (e.g., Capanu et al., 2013).

Latent Class Analysis

Model fit. Latent class analysis (LCA) was conducted at each testing wave to evaluate the model fit before the computing of estimation for the latent transition analysis (cf., Collins &

Lanza, 2010). Because LCA is an exploratory analysis, a series of models were computed varying the number of latent classes between one and seven (see Masyn, 2013, for a comprehensive review). A combination of statistical indicators and substantive theory was used to decide on the best fitting model. Statistical model comparisons included likelihood ratio tests: the Lo-Mendell-Rubin Test (LMR) and the Bootstrap Likelihood Ratio Test (BLRT). Both statistical procedures compared the improvement between neighboring class models (i.e., comparing models with three vs. four classes, and four vs. five, etc.) and provided *p*-values. *P*-values were used to determine if there was a statistically significant improvement in fit for the inclusion of one more latent class. A non-significant *p*-value for a K-class indicated that the previous K-class with a significant *p*-value fit the data better. The models with different numbers were compared using information criteria (i.e., Bayesian Information Criteria-BIC, Akaike Information Criteria-AIC, and Adjusted BIC). Lower values on these fit statistics indicated a better model fit. Among the information criterion measures, the Adjusted BIC and BIC indices are generally preferred, as is the BLRT for statistical model comparisons (Nylund et al., 2007; Nylund-Gibson & Choi, 2018). The SAS (Lanza et al., 2011) software was used to examine the manifest variables and determine the number of latent classes as well as perform the multilevel logistic modeling (see below). The Mplus (Muthén & Muthén, 2012) software was used to compute the likelihood ratio using the Lo-Mendell-Rubin Test (LMR) and the Bootstrap Likelihood Ratio Test (BLRT).

Nestedness. As noted, the data were nested within classrooms. At least 50 level-2 observations (classrooms in this case) are needed to assure that estimated parameters are unbiased (Maas & Hox, 2005). The number of clusters in our sample within each testing wave (i.e., 31 classes at Wave1, 23 at Wave 2, and 23 at Wave 3) was not sufficient to use in a multilevel latent class analysis (LCA). However, a two-level latent class analysis was conducted with the categorical variables at each testing wave. In this situation, it is recommended to consider entropy values when there is a multilevel structure to the data (see Kaplan & Keller, 2011, p. 53) because the preferred index (Bayesian Information Criterion, BIC) may underestimate the number of classes.

Latent Transition Analysis

Latent transition analysis (LTA) was used after the optimal number of latent classes had been selected at each time point. The analysis utilized the PROC LTA procedure in SAS version 9.4 (Lanza et al., 2011). A three-step method was used to determine the LTA model (e.g., Asparouhov & Muthén, 2014; Morin et al., 2018; Nylund-Gibson et al., 2014). The first step examined whether the same number of latent classes could be identified across three testing waves (i.e., configural similarity). Several authors have suggested fitting the covariate after determining latent models (Asparouhov & Muthén, 2014) so as not to alter the model. Thus, grade level (chronological age) was entered into the analysis when predicting latent classes in the subsequent logistic models. The second step integrated the three retained solutions (one at each time point) into a single LTA model, allowing for the estimation of transition probability solutions estimated across the three testing waves. The latent classes were constrained to be equal across the three test waves to examine the transitioning from one class to another over time. In applications of LTA, full measurement invariance is assumed for practical reasons, because it ensures that the number and structure of classes are the same across time and allows for a straightforward interpretation of transition probabilities. From these analyses, the following sets of parameters were estimated: latent class membership probabilities at Time 1, 2 and 3 (δ delta parameters), probabilities of transitions between latent classes over time (tau τ parameters), and item response probabilities conditional on latent class membership and time (ρ rho). Multiple LTA models were conducted to identify an adequate model fit of the data.

Multilevel Logistic Model

Cognitive measures. After determining the number of latent classes, a multilevel logistic model, via SAS PROC GLIMMIX software (SAS Institute, 2014), was used to analyze cognitive differences between the latent classes. This analysis determined those cognitive variables external to the classification measures that played a significant role in predicting latent class membership. Because we were not interested in the variance related to individual cognitive tasks, but what was common amongst the observed variables, as well as controlling for measurement error and enhanced reliability, latent measures served as predictors in the analysis. Previous analyses tested the categorization of the variables (i.e., working memory [WM], short-term memory [STM], naming speed, and inhibition) and provided a good fit to the data in a previous study (Author et al., 2019). Latent scores were computed by multiplying the z-score of the target variable by the standardized factor, loading weight based on the total sample (see Nunnally & Bernstein, 1994, p. 508, for calculation procedures). Latent variables were specified as indicators of speed (naming speed for numbers and letters), inhibition (random generation of numbers and letters), STM (Digit Forward Span, Word Span, and Phonetic Span), executive processing (Conceptual Span, Listening Span, Digit Sentence Span, Updating), and visual-spatial WM (Visual Matrix, Mapping & Directions). The correlations among the latent measures of cognition (grand mean centered across testing waves) are shown in Supplement Table 2.

As shown, measures were converted to a z-score with a mean of 0 and *SD* of 1, and the z-scores for Wave 2 and 3 (shown in Supplement Table 3) were based on the means and standard deviations of Wave 1. The means and standard deviations for cognitive measures (E-STM, S-

STM, E-Speed, S-speed, E-Inhibition, S-Inhibition, E-Executive WM, S-Executive WM, & visual-WM) as a function of latent class status and testing wave are shown in Supplement Table 3. Supplemental Table 5 shows ES comparisons among lantent classes at Wave 1, 2, and 3.

Model comparisons. Of interest in the logistic model was whether the log-odds of being identified within a particular latent class increased or decreased as a function of performance on specific cognitive measures. The reference group for the logistic regression was the latent class with the largest sample size (c.f., Hosmer & Lemeshow, 2000).

The equation for estimating the LCs for the unconditional (empty) model was:

$$\eta_{ij} = \beta_{0j} + \beta_{1j} X_{ij} \qquad \text{(Eq. 1)}$$

Equation 1 represented a simple level-1 model with one student-level predictor, where η_{ij} represented the log odds of reflecting a latent class other than an average achiever for student i in classroom j, β_{0j} is the intercept or the average log odds of being designated at risk in classroom j, ij is a student-level predictor for student i in classroom j, and β_{1j} represents the slope associated with X_{ij}, showing the relationship between the student-level variable and the log odds of not being designated at risk. It is important to note that unlike the hierarchical linear model used to analyze the total sample, this model has no error variance at level-1 (see Snijders & Bosker, 1999; pp. 225-227). As the effect of the student-level predictor was modeled as fixed or constant across classrooms, this was represented as a random intercept-only model.

$$\eta_{ij} = \gamma_{00} + \gamma_{10} X_{ij} + \gamma_{01} W_j + u_{0j} \qquad \text{(Eq. 2)}$$

The combined level-1 and level-2 model represented the log odds of being designated as a latent class at risk for student i in classroom j (η_{ij}) at a typical classroom ($\gamma 00$), at the student-level ($\gamma 10\ X$ij) and classroom-level predictor ($\gamma 01 W$j), as well as the classroom-level error [$u0$j, $u0$j ~ $N(0,\tau_{00})$].

The multilevel logistic modeling compared two conditional models (English only, Spanish only) and a full model (combined English and Spanish predictors) in determining the log-odds of identifying latent classes at the final testing wave. Because the full multilevel model had several predictors, and because the analysis included several variance components relative to simple linear regressions (Snijders & Bosker, 1999), alpha was set to .05. Bell et al. (2014) indicated in their Monte Carlo study that Type I error rates are not substantially inflated across multilevel models and conditions. The Bell et al study found across several multilevel design factors that the Type I bias was minimal and that 95% confidence interval coverage and Type I error rates tend to be conservative and well-controlled, even when modeling hierarchically structured data with smaller sample sizes. Also because we determined if the conditional models had a better fit than the unconditional model, deviance values (-2LL) and the Akaike's Information Criterion (AIC) were computed. The variance inflation factor (VIF) was computed on the fixed effects in the full model (see Table 9). All VIFs were less than 5, which is below Cohen et al.'s (2003) cut-off point, and therefore within an acceptable range.

Transparency and Openness

The materials described in the methods section are available from each of the authors, and are appropriately cited in the reference section. The data is based upon z-scores (converted from raw scores) and normed referenced scores for the latent transition analysis are included in a SAS (9.4) file entitled glimixyear123(rR) and is available from the Web site at APA's repository hosted by the Center for Science. The PROC LTA and LCA programs in the analysis were downloaded (free download) from www.methodology.psu.edu. The sample size (N=284) was determined by children participating in at least two testing waves during the three year period on vocabulary and achievement measures. Participants who participated in only one testing wave were not included in the analysis because the estimation of missing at random and/or the estimated value for the child's missing score could not be computed. A comparison of the two groups (retained vs. nonretained at Wave 2) on the Wave 1 manifest variables is shown in supplemental Table 6. The univariates yielded a significant advantage (.05/14, alpha=.004) on normative measures of Spanish calculation and Spanish word identification in favor of the nonretained group. No significant differences emerged related to gender, χ^2 (1,392)=.012, p=.91, whereas a significant effect emerged on measures of grade level, χ^2 (1,392)=23.27, $p < .001$. Non retained children were more frequently represented in grade 1 than retained children.

No outliers of children who participated in at least two testing waves were removed from the data analysis. Some measures were left out of the analysis (estimation, magnitude judgment) because we had no equivalent process measures in reading or they were not memory measures. The factor structure used to create the latent variables was reported in Swanson, Kong and Petcu (2019). The present work reported was not preregistered, but follows the stated purposes of the federal grant.

Results

Classification Variables

The means and standard deviations for the classification (manifest) variables for Waves 1, 2, and 3, as a function of the total sample and the three cohorts, are reported in Supplement Table 1. To capture performance changes across the testing waves, differences in the normative scores (Wave 3 normative scores minus Wave 1 normative scores) on the classification (manifest variables) measures were computed. As shown in Table 1, the negative difference scores showed that several measures (e.g., reading) decreased in normative values from Wave 1 to

Wave 3. To highlight declines we set a reference point of -5.00. As shown in bold, substantial declines (negative difference score > - 5.00) on normative tests in the total sample occurred on measures of English and Spanish mathematical problem solving and measures of Spanish reading comprehension. Declines on the normed manifest measures as a function of each cohort were as follows: English and Spanish math and Spanish comprehension for Cohort 1 (grades 1 to 3), Spanish vocabulary, math, and comprehension for Cohort 2 (grades 2 to 4), and English receptive vocabulary, Spanish computation, and Spanish comprehension for Cohort 3 (grades 3 to 5). The results also show a reduction in sample size at wave 3 for the CTRS measures. This was due to school closure and teachers were unavailable to fill out CTRS forms at the end of the school year.

Latent Class Analysis

Model Fit. Latent Class Model models were computed on the total sample (N=284) for Waves 1, 2, and 3. The results of our model testing at each testing wave are shown in Table 2. Of the indices reported in Table 2, the BIC and the adjusted sample BIC are generally preferred, as is the BLRT for statistical model comparisons (e.g., Nylund et al., 2007). In general, as shown, the BIC values were lower for the four latent class model (LC4) than for the five latent class model (LC5). The lowest adjusted BIC values emerged for the LC4 model. Both the LMR and BLRT yielded non-significant *p*-values for the solution of the latent class five (LC5) model, and significant *p*-values for the solutions of three latent class models, indicating that the LC4 model provided an excellent fit to the data. The LC5 model did not represent an improvement of the LC4 model, and the LC4 model was an improvement to the LC3 model.

As shown in Table 2, the entropy values for the four latent class models were .85, .84, and .87 for Waves 1, 2, and 3, respectively. These three values are considered acceptable

(Nylund et al., 2007). The null hypothesis that data are missing completely at random (MCAR) was tested for each model. For example, the log-likelihood divided by the degrees of freedom for the four latent class model at wave 1 (-1901.88/16324) was not significant (*p*=~1) suggesting the data were missing at random. Additional analyses suggested that the missing data were at random across all the latent class models.

Table 3 shows the estimated proportion of samples (based on Delta estimates) for the four-class model at each testing wave. Except for Latent class 1, the four-class model represented a relatively large number of participants (e.g., proportionally was greater than 5%) within each latent class (see Delta estimates). Consistent with Nylund-Gibson and Choi (2018) we also inspected the "elbow" of the point of "diminishing returns" in determining the model fit (e.g., small decreases in the IC (information criteria) for each additional latent class) (p. 443-444). Given the findings related to each testing wave, we now address our first question.

1. Can a stable latent class of EL children be identified with learning disabilities among children who vary in language acquisition?

Transition Analysis

To answer the first question, transition probabilities (tau τ parameters) were estimated for the total sample to identify the rate of change or stability over time for the latent class groups. The transition probabilities are shown in Table 4. Although the four latent class model was supported at Waves 1, 2 and 3 separately (establishing configurable similarity), we determined if the four latent class model was a good fit for the latent transition (LTA) model. A measurement variance model was computed that yielded BIC values of 8867.91, 8686.01, and 8707.38 for the LC3, LC4, and LC5 models in LTA, respectively. Because of the elbow between the LC3 and LC4 models, one could argue that the LC3 model provided a more parsimonious fit than the LC4 model. However, when we examined moving from the LC4 to the LC3 model, we noted that some classes (LC2 and LC4) were being combined, while the classes in the middle of the overall data (.00) distribution remained almost unchanged. This is not an unusual finding. As stated by Ryoo and colleagues (2018), "model parsimony is not necessarily the goal of LTA. The goal of LTA, particularly when it is used in an exploratory vein, is to describe the data by identifying classes within the population of observations in the data (p.31)." Thus, our selection of the LC4 model provided a more nuanced description of the data.

Latent class stability. To further analyze class stability, however, it is useful to constrain each element of the matrix of ρ parameters at Wave 1 to be equal to its corresponding element at Waves 2 and 3, which has the effect of imposing measurement invariance across time (Kam et al., 2016; Nylund-Gibson et al., 2014). This analysis forced the change process to be stationary (i.e., forced individuals transitioning between classes to have the same probability level of change across the three-time points). This comparison allowed us to check whether the latent classes have the same meaning across the three testing waves and three cohorts. The objective of researching measurement invariance is to find the lowest level of "inequivalence" possible that fits the data well.

As shown in Table 4, the transition probabilities of the latent classes (τ parameters) for the transitional model reflected membership in the same latent class model at three consecutive measurement times. Transition probabilities off the diagonal reflected the likelihood of one latent class status group at Wave 1 transitioning to a different latent class status group at Wave 2 and a latent class group at Wave 2 transitioning to Wave 3. For example, for the total sample, the LC1 status group at Wave 1 yielded a 100% estimated chance of maintaining their status at Wave 2 and an estimated chance of 86% at wave 3. In contrast, the LC2 status group in Wave 1 had only a 49% chance of being in the same group in Wave 2, suggesting that the item estimates were highly unstable. In contrast, the LC3 status group and the LC4 status group estimates were fairly stable from Wave 1 to Wave 2 (78% and 93%, respectively) and from waves 2 to 3 (93% and 81%, respectively).

To unpack the transitions that may be occurring across cohorts, the findings were analyzed by dividing the analysis into the grade-level cohorts. As shown in Table 4, the LC2 status group classification was the least stable for Cohort 2 (grades 2 to 4) at wave 1 (26%) with approximately 59% of the sample falling into the LC4 status group at Wave 2. This latent class was also unstable when transitioning at wave 2 from grade 3 to grade 4 (56% probability)

Labeling of latent classes. The demographic information and the means and standard deviations on the manifest variables and cognitive variables as a function of the four latent class status groups are reported in Supplement Table 3. A comparison of gender representation among the latent classes was not significant, χ^2 (3, N=284), 1.67, p=.64. However, as shown in Supplement Table 3, there was a clear pattern within cohorts that the lower grades were most likely assigned to the bilingual average achievers' group (LC2, i.e., scores in Spanish approximated those in English), whereas English dominant average achievers were most frequently represented in the upper grades, χ^2 (6, N=284),= 44.25. 41.16, 44.11, *p*s < .001, for grade level within cohorts 1,2 and 3, respectively. This pattern is consistent with the increasing emphasis on English instruction in the upper grades.

Table 5 highlights in **bold** the mean levels of performance at or below standard scores of 85 at Testing Wave 3, as well as highlights difference scores (Wave 3 minus Wave 1) showing a reduction of at least 5 standard score points. To facilitate the labeling of the latent classes, the frequency of low mean scores was first considered. As previously indicated, there were four measures of math (English and Spanish problem solving and calculation) and 4 measures of reading (English and Spanish measures of word identification and comprehension) at each testing wave. The number of **mean scores** for the eight achievement measures in Table 5 that were **below** 85 as a function of the latent class were: 6, 0, 4, and 1 for LC1, LC2, LC3, and LC4, respectively. The number of mean **difference scores** greater than -5 was 6, 7, 5, and 3 for LC1, LC2, LC3, and LC4, respectively. Taken together, we infer that LC 1 and LC3 showed the greatest risk for achievement difficulties relative to the other two groups. As shown, the level of performance for the LC 1 group yielded low mean scores (< 85 normative standard scores) across both math and reading measures. Thus, we label LC 1 as a comorbid learning disability group. In contrast, LC 3 manifested difficulties that were primarily isolated to high-order tasks, such as mathematical problem-solving and reading comprehension. Thus, this group was labeled as experiencing learning disabilities on problem-solving tasks. That is, their difficulties were isolated to high order tasks (mathematical problem solving and reading comprehension) rather than low order tasks, such as calculation and word identification.

In terms of bilingual status, the LC 2 group at wave 3 showed achievement and vocabulary in the average range on both English and Spanish measures. Thus, we label LC 2 as balanced bilingual average achievers. When compared to the LC2 group, LC 4 showed relatively low performance on Spanish measures relative to the English measures and therefore this latent class was labeled as unbalanced bilingual average achievers.

Magnitude differences. To further facilitate the labeling of the latent classes, the magnitude of differences in the manifest variables as a function of latent class status groups was computed. Using Cohen's (1988) criteria for medium (.50) to large (.80) effect sizes (ESs), ESs at or greater than .80 are shown in bold for performance in Table 6 for Wave 3 (Supplement Table 4 reports effect sizes at Wave 1 and Wave 2).

As shown in Table 6, an advantage was found for the LC2 group when compared to the LC1 group across the majority of achievement measures (all ES >.80). A performance advantage was also found for the LC2 group when compared to the LC3 group on all math measures. Based on effect size differences, the LC1 and LC3 status groups were considered at risk for learning disabilities.

The performance of average achievers (LC2 and LC4) was also compared. A higher level of performance occurred in the LC2 status group (balanced bilingual-average achievers) when compared to the LC4 status group (unbalanced bilingual-average achievers) at Wave 3 on measures of Spanish vocabulary, Spanish math problem solving, and Spanish reading comprehension. Thereby further suggesting that the LC 2 reflected a balanced bilingual achieving group.

Item probabilities. Item probabilities (rho estimates) based on the four latent class transition models are shown in Table 7. These rho estimates reflected the latent class abilities of the given item-response, conditional on the given latent-class membership. The probabilities (rho estimates) shown in Table 7 reflect performance *below* the cut-off threshold of the 25th percentile (90 standard score) on the manifest variables. To facilitate discussion, and because there is no set standard for determining meaningful probabilities, item latent class abilities above 60% were selected, and these values are shown in bold. That is, probabilities above .60 indicated "at-risk" status for that particular manifest variable. The results show that all of the parameters of vocabulary, reading, and math for the LC1 status group were above .60. The only exception was measures of classroom attention. All probabilities for the LC2 group were less than .50 and therefore were consistent with our earlier classifications as a balanced bilingual average achiever group. The LC3 status group showed high risk on measures of English vocabulary, English problem solving, and English reading comprehension; whereas the LC4 status group showed low Spanish reading comprehension and vocabulary.

In summary, the previous analyses showed a distinct latent and stable class of children with comorbid achievement difficulties (LC 1) and difficulties in math problem solving/reading comprehension (LC 3) emerged within the data set. The results also suggest variations among children in terms of language acquisition. Two latent classes related to variations in language acquisition emerged related to average achievement. One latent class included average achievers scoring in the average range on both Spanish and English vocabulary (LC 2) measures, whereas another latent class of average achievers excelled in their second language relative to their first language (LC 4). We now address the second question of the study.

2. **Do measures of growth and the level of cognitive performance predict the latent class status?**

Multilevel Logistic Model

A multilevel logistic model determined those cognitive variables that uniquely predicted the log-odds of LC status groups at Wave 3. Because it is recommended to use the largest sample as a reference group (c.f., Hosmer & Lemeshow, 2000), the three latent classes (LC1, LC2, LC) were compared to LC4 (unbalanced bilingual average achievers). The estimates for the multilevel logistic unconditional model predicting the log odds of being classified in one of the latent classes at wave 3, when compared to the LC4 status group, as a function of variables external to the classification (cognitive variables), are shown in Table 9. Table 9 shows the three intercept values for the LC1 (low achievers-comorbid), LC2 (balanced-average achievers), and

LC3 (at risk for math/reading learning disabilities) status groups when compared to the LC4 status group.

Empty (unconditional) Model. The empty model shown in Table 8 was assumed to have no error at level-1 (Snijders & Bosker, 1999). The level-1 residual follows a logistic distribution with a mean of 0 and a variance of 3.29 (Snijders & Bosker, 1999, p. 227). The cross-sectional intercept for clusters (classroom) at Waves 1, 2, and 3 would not converge because of too many random effects on the data. Therefore, level two-variance was computed from testing Wave 2 because it yielded the largest variance relative to the other cluster effects at waves 1 and 3. The estimated intercept (level 2) variance at wave 2 was .32, .40, and .67 for the three comparisons (LC 1 vs. LC4; LC2 vs. LC4; LC3 vs. LC4, respectively). The intraclass correlation was computed as .30 (.32 + .40 + .67 /.32 + .40 + .67 + 3.29), suggesting that approximately 30% of the variability was accounted for by children nested in classrooms at Wave 2, leaving approximately 70% of the variability to be accounted for by the latent measures (or other unknown factors).

Language-specific conditional models. Table 8 also shows two conditional multilevel logistic models that simultaneously entered the cognitive measures to predict latent class status at wave 3. All predictor measures (intercepts and slopes) were grand mean-centered. The first model considered the contribution of English cognitive measures and the second model considered the contribution of Spanish cognitive measures in predicting latent class status. As shown in Table 8, the estimated level two-variance increased from the empty model to the conditional models. In contrast to hierarchical linear models with normal distributions and increased level-two variance that would suggest a misfit of the data, models with dichotomous data tend to increase level-two variance when adding level 1 variables (see Snijiders & Bosker, 1999, p. 217, for discussion). As stated by Snijders and Bosker " adding level 1 variables with strong effects will increase estimated level-two variances, and to make the regression coefficients of already included variables, if these are uncorrelated with newly included variables, larger in absolute size" (p.217). For interpretation of significant level 1 parameters, a significant positive beta weight suggested that the log-odds of identifying the two latent classes (e.g., LC 1 vs. LC4) increased related to increased performance on that measure. Significant negative beta weights suggested the log-odds of identifying the groups decreased based on increased performance for that measure.

As shown in Table 8, significant English predictors of the LC1 vs. LC4 intercept (log odds) were the level of performance on the inhibition (b=-.60) and WM (b=-.85) measures. To interpret these findings, as well as other comparisons, it is important to note that the significant coefficient of -.60 for inhibition implied that a one-unit change in inhibition would result in a -.60 unit change in the log odds of identifying the two groups. Table 8 also shows that the level of STM and WM performance and the visual-spatial WM slope was significantly related to the log-odds of the LC3 vs. LC4 intercept. No significant cognitive parameters emerged predicting the log-odds of the LC2 vs. LC4 comparison.

The results for the Spanish predictors are shown at the bottom of Table 8. Significant Spanish predictors of the LC1 vs. LC4 log-odds were intercept measures of STM, WM, and visual WM and the growth measure of STM. The level of WM and naming speed performance predicted the LC2 vs. LC4 log-odds. No significant parameters predicted the LC3 vs. LC4 log-odds. As shown at the bottom of Table 8, the AIC values are reported and allowed for the comparisons of nonnested models (Hox, 2010, pp. 45-50). When comparing the English and

Spanish models, the better fit to the data based on the lowest AIC values was the English-only model.

Full conditional model. Table 9 shows the results when measures of English and Spanish are simultaneously entered into the model. The log-odds comparing the LC1 vs. LC4 were predicted by the level of performance on English measures of STM and WM and growth on measures of Spanish STM and English WM. Likewise, the log-odds for predicting the LC2 vs. LC4 comparison were significantly related to the intercept measures of English STM, English naming speed and the growth parameter related to English WM. Spanish measures that significantly predicted the log-odds of the LC2 vs. LC4 comparison were measures of naming speed and WM. Finally, the log-odds for predicting the LC3 vs. L4 comparison were significantly related to the measures of English STM, naming speed, and WM and Spanish measures of naming speed. Growth on Spanish measures of WM also predicted the log-odds of the LC3 vs. LC4 comparison.

3. Do Spanish cognitive processes enhance or operate independently of English cognitive predictions of latent class status?

Taken together, the entry of the Spanish measures into the Full Multi-level logistic regression model yielded three important findings. First, based on a total of eight English measures (4 for intercept and 4 for growth), the results indicated that 50% of these measures in the LC 1 vs. LC4 comparison, 62% in the LC2 vs LC4 comparison, and 62% in the LC3 vs. LC4 showed an increase in beta weights when compared to the English only model. The largest increase in magnitude for English measures was captured by the level of performance on STM (increase -.53) for the LC1 vs. LC4 comparison, naming speed (increase= .52) for the LC2 v. LC4, and naming speed (increase= .71) for the LC3 vs. LC 4 comparison. All comparisons in the

Full model showed the largest increase in the growth parameters for WM when compared to the other English-only model for growth parameters.

Second, Spanish measures contributed significant variance to the log-odds of balanced bilingual and unbalanced bilingual comparisons. The results showed that no significant predictors emerged related to the log-odds of identifying the LC2 relative to the LC4 reference group in the English-only model. However, significant predictors emerged on measures of STM, naming speed, and WM in the Full Model. Finally, the Full Model provided a better fit to the data than the English-only model. The statistical contribution of Spanish measures in the Full model relative to the English-only model was determined by comparing the deviance values (-2LL). There was a significant difference between the Full and English-only model, χ^2 (df=8; 1153.55 – 836.30)= 317.25, $p < .001$.

In general, the results show that the beta-weights for L2 cognitive performance increased when L1 cognitive measures were entered into the analysis. Although the contribution of Spanish measures was most apparent in the LC2 vs. LC4 comparisons, all comparisons showed that entry of Spanish measures enhanced the significant role of WM growth in predicting latent class status. Thus, the significant predictions of English cognitive measures on latent class status were not independent of performance on Spanish cognitive measures.

Discussion

One purpose of this study was to determine if a stable class of EL children at risk for learning disabilities could be identified from other EL children who were average achievers but varied in first and second language proficiency. Three important findings emerged. First, two stable latent class groups at risk for learning disabilities were identified. The LC1 (low math and reading) and LC3 status group (low mathematical problem-solving and reading comprehension) showed average intelligence and attention, but stable difficulties in performance across both English and Spanish measures of achievement. Second, cognitive variables uniquely predict latent class status. Those cognitive measures in the Full Model that played a key role in latent class predictions at the intercept level were measures of STM and naming speed. The only growth measure that predicted latent classes in the Full Model was WM.

Finally, cognitive performance in L1 substantially influenced the magnitude of the predictions of L2 cognitive performance on the log-odds of identifying latent classes. The results clearly showed that including Spanish measures in the analysis increased the log-odds of identifying balanced bilingual average achievers from unbalanced average achievers on measures of naming speed, STM, and WM. In addition, the log-odds of identifying both average and learning disability groups as a function of growth (i.e., working memory) were significantly enhanced by including Spanish measures in the analysis. Taken together, these results suggest differences between the latent classes are heightened by difficulties or strengths in their first language. Thus, cognitive difficulties that potentially underlie the child's second language do not operate independently of their first language. These results support the notion that there is an interchange between the two language systems that underlie latent class status. Whether this interchange reflects accenting the key processes in the second language by inhibiting other processes and/or providing L1 resources to L2 performance, is discussed below.

We now consider three questions that directed this study.

1. **Can a stable latent class of EL children be identified with learning disabilities among children who vary in language acquisition?**

The results answer this question in the affirmative. The results showed that latent class groups (LC1 and LC 3) emerged for children at risk in reading and/or math in both language systems, and these latent classes had a low probability of transitioning into average achieving latent classes. As shown in our analysis (see Table 4), the latent class membership transition probabilities for the latent class of EL children at risk for learning disabilities in both reading and math (LC1) was approximately 100% from Wave 1 to Wave 2 and approximately 86% from Wave 2 to Wave 3. Likewise, the latent class membership transition probabilities for the latent class of EL children at risk for learning disabilities primarily on high-order tasks (math problem solving and reading comprehension) (LC3) was approximately 78% from Wave 1 to Wave 2 and approximately 93% from Wave 2 to Wave 3.

Further analysis of probabilities within each cohort suggested the latent classes for children at risk for learning disabilities were stable. Latent class 1 (comorbid low achievers) classification was stable across all cohorts and testing waves (100%). Latent class 3 showed some instability in Cohort 2 transitioning from wave 1 to wave 2 (65% maintaining classification), but the stability increased transitioning from Wave 2 to 3 (95% maintained the classification).

Some comment is necessary on the latent class of average achievers found the least stable (LC2). As shown in Table 3, the balanced bilingual achiever (LC2) showed low stability across all cohorts. The results further suggest that a large number of these children transition into the unbalanced (English dominant) bilingual average achiever group. This finding coincides with a decrease in Spanish performance from wave 1 to wave 3. This finding coincides were our observation that the language of instruction in the older grades placed a heavier emphasis on English instruction, rather than allowing for access to information from both the English and Spanish language systems. As English learners continue to develop competencies in their second language, there is often a decline in their L1, or Spanish language skills (also see Gottardo et al., 2008; for the difference in instructional strategies across grades). When the Spanish language is no longer emphasized or supported in the classroom, the opportunities for maintaining and continuing to build L1 diminish. Indirect evidence for this outcome was the high incidence of English dominant achievers in Wave 1 (LC2) transitioning to low Spanish proficiency (LC4) at the second or third testing wave.

The results also suggest there may be late-emerging difficulties on measures of mathematical problem solving and reading comprehension. As shown in Table 4, thirteen percent (13%) of children labeled as balanced bilingual average achievers at wave 1 (referred to as LC 2) were labeled at risk for learning disabilities in mathematical problem solving and reading comprehension (LC 3) at wave 2 and 14% at wave 3. At wave 1, as a group, these children (LC2) yielded English and Spanish vocabulary, math, and reading scores in the average range (> 85 mean standard scores). However, the incidence of late-emerging risk among some of these children occurred across all cohorts. For example, when transitioning from Wave 1 to wave 2 and wave 2 to wave 3, the probabilities by cohort were: 13% and 20% for Cohort 1 (grade 1 to grade 2, grade 2 to grade 3), 14% and 0% for Cohort 2 (grade 2 to grade 3, grade 3 to grade 4), and 10% and 5% for Cohort 3 (grade 3 to grade 4, grade 4 to grade 5), respectively. Thus, by using both language systems that focused on additional performance measures of risk beyond reading and receptive language, we were able to identify several children who were average achievers at wave 1, but exhibited risk status by wave 3.

Finding children with late-emerging learning disabilities, especially as it applies to reading comprehension, is consistent with the literature (Barber et al., 2022; Cho et al., 2019; Miciak et al. 2022; Taboada et al., 2022). Finding latent emergent children among EL children is also consistent with our earlier work with monolingual elementary identifying a latent class of students with low achievement in mathematical problem-solving and reading comprehension, but average performance in areas of word identification and fluid intelligence (Swanson et al., 2018). In addition, the link between comprehension and word problem solving is consistent with the notion mathematical word problems are a form of text and decoding and the comprehension of text draws upon a common system (cf., Fuchs et al., 2018; Iglesias-Sarmiento et al., 2015). Although the cognitive variables that link problem-solving and reading comprehension are still under study, some studies suggest that numeracy may mediate both processes (e.g., Spencer et al. 2022), others mathematic knowledge (Papura et al., 2017), whereas others see working memory (e.g., updating, Cornoldi et al . 2012) as playing a mediating role. Our data cannot address this issue. However, as shown in Table 5, the difference scores on both English and Spanish vocabulary were small, whereas large difference scores occurred on the majority of reading and math tasks. Thus, one could infer that late-emerging difficulties for some EL children may be related to cross-language transfer on reading and math tasks.

These aforementioned cross-language difficulties for some children with late-emerging learning disabilities appear related to inefficiencies in their first language. That is, the multilevel logistic modeling showed that the log-odds of identifying balanced average achieving children relative to unbalanced average achievers occurred primarily in the first language (L1) system (see Table 8). The largest beta weights occurred on Spanish measures of naming speed and working memory (see Table 8).

Performance on first language measures also plays an important role in predicting stable learning disabilities status. On the surface, the results appear to suggest that L1 processes play a more important role in separating children that vary in language acquisition than children with stable learning disabilities. That is, the significant effects of entry L1 measures into the full logistic model were more frequent for average achievers than children with LD, suggesting that L1 processes play a more important role in language acquisition among average achievers than children at risk for learning disabilities. This inference, however, has to be reconciled with two findings. First, 50% of the English measures beta weights in the LC1 group and 62% of the English beta weights in the LC3 group were increased when Spanish measures were entered into the model. Second, as shown in supplement Table 5, moderate to large effect sizes occurred on the majority of Spanish cognitive measures across all testing waves for the LC1 (comorbid learning disabilities) vs LC2 (balance average achievers) comparison. In addition, increased growth in Spanish WM decreases the log-odd of identifying children with high-order difficulties (LC3) when compared to the unbalanced bilingual average achievers (see Table 9). Thus, we infer that the identification of children with learning disabilities is related to cognitive processing in the L1 system. In general, we infer that both stable and late-emerging learning disability status is best identified by the contribution of L1 measures to L2 cognitive performance.

2. Do measures of growth and the level of cognitive performance predict latent class status?

The results suggest that cognitive measures do uniquely predict latent classes. Cognitive measures, such as the levels of STM performance and WM slopes, were associated with latent class membership. However, results of the full multilevel logistic regression model do not support our hypothesis that latent classes related to language acquisition are primarily predicted by measures related to the phonological loop (STM, naming speed) and latent classes related to learning disabilities are primarily predicted by measures of the executive component of WM. Both the phonological loop (STM, naming speed) and the executive system (WM) predicted the language acquisition (LC2 & LC4) and learning disabilities (LC1 and LC3) latent classes. For example, as shown in Table 9, the odds of significantly identifying children with comorbid learning disabilities related to unbalanced bilingual average achievers was related to performance on measures of English STM and growth in English WM. Performance on measures of English STM, English and Spanish naming speed, and Spanish WM predicted the log odds of significantly identifying balanced bilingual average achievers from unbalanced bilingual average achievers. Likewise, the log odds of significantly identifying children with learning disabilities in math problem solving/ reading comprehension from unbalanced bilingual average achievers was on measures of English measures of STM, English and Spanish naming speed, and English WM. The only growth parameter that predicted the log-odds of learning disability status (LC1 and LC3) or language acquisition status (LC2) relative to the referent group (LC4) were measures of WM.

3. Do Spanish cognitive processes enhance or operate independently of English cognitive predictions of latent class status?

As shown when comparing the two conditional models in Tables 8 and 9, entry of Spanish cognitive measures enhanced predictions of L2 measures (by increasing the magnitude of the beta-weights) in determining the log odds of identifying the latent classes. Support for this assumption was found by comparing the beta-weights between the two models as well as a comparison of the deviance values. The importance of entering Spanish measures is particularly salient in the LC2 vs. LC 4 comparison. None of the cognitive measures were significant when only English tasks were entered into the analysis, whereas the importance of memory measures emerged in the full model.

In general, the literature has been unclear as to the degree to which EL children experience cognitive benefits related to having access to both language systems. The results reported in Tables 8 and 9 show that L1 cognitive measures enhance the predictability of English measures on LC2 vs, LC4 status. For example, as shown in Table 9, the contribution of cognitive measures to the log-odds comparing LC2 with LC4 was significant only when Spanish measures were entered into the analysis. These findings suggest that L1 cognitive performance (in this case Spanish naming speed, STM WM, and inhibition) does influence the contribution of L2 cognitive performance to identifying latent classes LC2 from LC4. Even though few English cognitive measures were significant in the latent class comparison, the entry of Spanish measures boosted the beta weights of several English measures. These findings are important because previous research has been unclear as to whether the relationship between learning disabilities in reading and/or math in EL children in the first and second language reflects cognitive costs related to the transferring from one language system to another, or if the relationship is merely an artifact of L1 or L2 proficiency (e.g., Choi et al., 2018; Kempert et al., 2011).

Although the aforementioned findings may bring into light the positive influence of L1 processes on L2 cognition, a clarification should be made of our goals for including Spanish measures in our analysis of latent class status. Although the limitations of using repressor variables as well as distinguishing such an approach from mediation analysis have been discussed in the literature (e.g., Kim, 2019); our goal was to make an accurate prediction of LC status as a function of cognitive performance. On a practical level, it has been unclear as to whether assessing a child in his/her first language offers any additional information about a cognitive processing deficit that can be obtained by just testing a child in the second language. This is especially true if the language of instruction is geared primarily toward the child's second language.

Our findings suggest, however, that there is a risk in limiting the predictability of latent class status in EL children by excluding L1 cognitive variables. An incomplete set of independent variables may not only underestimate regression coefficients but in some instances, may fail to reject the null hypothesis when it is false (Type II error). Although including irrelevant variables in a model can contribute to loss of degrees of freedom and/or multicollinearity, several studies suggest those variables will not affect the predictive power of the model (e.g., Cohen et al., 2003; Kim 2019; Tzelgov & Henik, 1991). Hence, the risk of longitudinal studies excluding Spanish variables outweighs the risk of including such variables in the analysis. Our study does bring into question several longitudinal studies that assess risk for academic difficulties only in the child's second language (English).

Limitations

There are at least four limitations in the study (see Swanson et al., 2021 for further limitations at the earlier testing waves). First, our sampling reflected children who were sequential bilinguals (L2 follows L1 development) and therefore may not reflect bilingualism when two languages are learned simultaneously (e.g., Sabourin & Vīnerte, 2015). Also, our geographic region (e.g., studies conducted in the southwest U.S. versus Canada) may limit the generalizability of the findings. Second, we used normative measures, and therefore it is important to note that some of the standardized assessments do not report the language background of their norming sample or are based primarily on a monolingual English-speaking population. When comparing second language (L2) speakers' performance using the standardized scores, it is relative to others who speak a single language. Given that there is research showing that Spanish-English bilingual children may show different developmental trajectories in both languages (e.g., Hoff, 2017), it is possible to observe 'lower' standard scores as children age, reflecting the slower growth.

Third, because the majority of the sample (> 95%) was participating in the Federal free and reduced lunch program, we did not have enough variance to assess the role of SES in the performance outcomes. Although participating in a Federal Free lunch program may be insensitive to various parameters when assessing SES (Harwell & LeBeau, 2010), participation in the program is based on low family income. Greater resources in the home (e.g., access to books and homework assistance) would have a positive influence on math and literacy growth in children. This issue could not be addressed within our data set.

Fourth, using the 25th percentile as an “a priori” cut-off score for determining risk for learning disabilities is a potential limitation because the same or other groups may have emerged with other cut-off scores. Thus, we have not shown that the identification validates a specific cut-off point; rather the results suggested that the measures were able to identify subgroups to the cut-off for which they were applied (Swanson et al., 2016, 2018). Although the 25th percentile or below across multiple years has been used as an “a priori cut-off point” to identify children at risk for learning disabilities, the issue as to whether other cut-off scores yield similar groups of EL children at risk and/or discrete or identifiable groups has not been established.

Finally, our results must be considered in light of the limitations inherent in the LTA model. Model selection in LCA and LTA is a significant challenge in mixture models (e.g., Dziak et al., 2014) because little is known about statistical power in LCA and LTA. Although several model selection indices of LTA are available, model selection is best conducted when a great deal of attention is paid to model interpretability (Lanza et al., 2010).

Summary

In summary, this study yielded three important findings. First, latent classifications of children at risk for learning disabilities could be identified among a sample of ELs. Two latent classes at risk for learning disabilities emerged across the three testing waves. Second, children at risk for learning disabilities could be separated from children who varied in language acquisition on cognitive measures. Finally, the log odds of identifying latent classes at the third testing wave are related to the level of performance and growth (slopes) on measures related to the phonological loop (STM and naming speed) and the executive component of WM. These outcomes of identifying children with potential learning disabilities were further enhanced by taking into consideration performance on Spanish measures. Thus, the results suggested that the probability of risk for learning disabilities versus language acquisition can be tied to cognitive performance in their first language.

References

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders, 5th edition*. Washington, DC: American Psychiatric Association.

Artiles, A. J., Rueda, R., Salazar, J. J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children, 71*(3), 283-300. doi: 10.1177/001440290507100305

Ashkenazi, S., Rosenberg-Lee, M., Metcalfe, A. W. S., Swigart, A. G., & Menon, V. (2013). Visuo–spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia, 51*(11), 2305-2317. doi:/10.1016/j.neuropsychologia.2013.06.031

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using M-Plus. *Structural Equation Modeling, 21*(3), 329-341. doi:10.1080/10705511.2014.915181

Baddeley, A. (2007). *Working memory, thought, and action*. New York, NY, US: Oxford University Press. doi:10.1093/acprof:oso/9780198528012.001.0001

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1-29. doi: /10.1146/annurev

Baddeley, A. D., & Logie, R. H. (1999). The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance* and executive control (pp. 28-61). Cambridge, U.K.: Cambridge University Press.

Barber, A. T., Klauda, S. L., Wang, W., Cartwright, K. B., & Cutting, L. E. (2022). Emergent bilinguals with specific reading comprehension deficits: A comparative and longitudinal analysis. *Journal of Learning Disabilities, 55*(1), 43-57. doi:https://doi.org/10.1177/0022219420983247

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*(1), 71-92. https://doi.org/10.1037/0096-3445.132.1.71

Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Kromrey, J. D., & Ferron, J. M. (2014). How low can you go? An investigation of the influence of sample size and model complexity on point and interval estimates in two-level linear models. *Methodology: European Journal of Research Method for the Behavioral and Social Sciences, 10*(1), 1-11. https://doi.org/10.1027/1614-2241/a000062

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology*, *65*(4), 229-235. https://doi.org/10.1037/a0025406

Bialystok, E., & Feng, X. (2009). Language proficiency and executive control in proactive interference: Evidence from monolingual and bilingual children and adults. *Brain and Language, 109*(2-3), 93–100. 10.1016/j.bandl.2008.09.001

Boada, R., Willcutt, E. G., & Pennington, B. F. (2012). Understanding the comorbidity between dyslexia and attention-deficit/hyperactivity disorder. *Topics in Language Disorders, 32*(3), 264-284. doi:10.1097/TLD.0b013e31826203ac

Bonifacci, P., Giombini, L., Bellocchi, S., & Contento, S. (2011). Speed of processing, anticipation, inhibition and working memory in bilinguals. *Developmental Science, 14*(2), 256-269.

Brandenburg, J., Klesczewski, J., Schuchardt, K., Fischbach, A., Büttner, G., & Hasselhorn, M. (2017). Phonological processing in children with specific reading disorder versus typical learners: Factor structure and measurement invariance in a transparent orthography. *Journal of Educational Psychology, 109*(5), 709-726. doi:10.1037/edu0000162

Branum-Martin, L., Fletcher, J. M., & Stuebing, K. K. (2013). Classification and identification of reading and math disabilities: The special case of comorbidity. *Journal of Learning Disabilities, 46*(6), 490-499.

Brownell, K. (2001). *Expressive One-Word Picture Vocabulary Test* (3rd Edition). New York: Academic Therapy Publications.

Capanu, M., Goenen, M., Begg, C. B. (2013). An assessment of estimation methods for generalized linear models with binary Data. *Statistical Medicine. 2013 Nov 20;* 32(26): 10.1002/sim.5866. doi: 10.1002/sim.5866

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology, 104*(1), 166-181. doi: 10.1037/a0025323

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading, 6*, 369-400. doi:10.1207/S1532799XSSR0604_04

Chiappe, P., & Siegel, L. S. (2006). A longitudinal study of reading development of Canadian children from diverse linguistic backgrounds. *The Elementary School Journal, 107*(2), 135-152. doi:10.1086/510652

Child, A. E., Cirino, P. T., Fletcher, J. M., Willcutt, E. G., & Fuchs, L. S. (2019). A cognitive dimensional approach to understanding shared and unique contributions to reading, math, and attention skills. *Journal of Learning Disabilities, 52*(1), 15-30. doi:10.1177/0022219418775115

Cirino, P. T., Fuchs, L. S., Elias, J. T., Powell, S. R., & Schumacher, R. F. (2015). Cognitive and mathematical profiles for different forms of learning difficulties. *Journal of Learning Disabilities, 48*(2), 156-175. doi:10.1177/0022219413494239

Cho, E., Capin, P., Roberts, G., Roberts, G. J., & Vaughn, S. (2019). Examining sources and mechanisms of reading comprehension difficulties: Comparing English learners and non-English learners within the simple view of reading. *Journal of Educational Psychology, 111*(6), 982-1000. doi:10.1037/edu0000332

Cho, E., Fuchs, L. S., Seethaler, P. M., Fuchs, D., & Compton, D. L. (2020). Dynamic assessment for identifying Spanish-speaking English learners' risk for mathematics disabilities: Does language of administration matter? *Journal of Learning Disabilities, 53*(5), 380-398. doi:10.1177/0022219419898887

Choi, J. Y., Jeon, S., & Lippard, C. (2018). Dual language learning, inhibitory control, and math achievement in head start and kindergarten. *Early Childhood Research Quarterly, 42*, 66-78. doi.org/10.1016/j.ecresq.2017.09.001

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences, 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. New Jersey: Wiley and Sons.

Colom, R., Shih, P. C., Flores-Mendoza, C., & Quiroga, M. Á. (2006). The real relationship between short-term memory and working memory. *Memory, 14*(7), 804-813. https://doi.org/10.1080/09658210600680020

Conger, A. J. (1974). A revised definition for suppressor variables: A guide to their identification and interpretation. *Educational and Psychological Measurement, 34*(1), 35-46. doi:10.1177/001316447403400105

Conners, C. K. (1997). *Conners' Rating Scales-Revised: Technical manual.* North Tonawanda, NY: Multi-Health Systems.

Cornoldi, C., Drusi, S., Tencati, C., Giofrè, D., & Mirandola, C. (2012). Problem solving and working memory updating difficulties in a group of poor comprehenders. *Journal of Cognitive Education and Psychology, 11*(1), 39-44. doi:https://doi.org/10.1891/1945-8959.11.1.39

Cowan, N. (2014). Working memory underpins cognitive development, learning, and education. *Educational Psychology Review, 26*(2), 197-223. doi:10.1007/s10648-013-9246-y

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49,* 222-251. doi:10.3102/00346543049002222

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466. https://doi.org/10.1037

David, C. V. (2012). Working memory deficits in math learning difficulties: A meta-analysis. *International Journal of Developmental Disabilities, 58*(2), 67-84. doi:/10.1179/2047387711Y.0000000007 de Abreu, P. M. J.. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19*(5), 529-537.doi:10.1080/09658211.2011.590504

de Abreu, P. M. J., & Gathercole, S. E. (2012). Executive and phonological processes in second language acquisition. *Journal of Educational Psychology*, *104*(4), 974-986. doi:10.1037/a0028390

Dunn, L. M., & Dunn, L. M. (2007). *The Peabody Picture Vocabulary Test-4.* NY: Pearson.

Dunn, L. M., Lugo, D. E., Padilla, E. R., & Dunn, L. M. (1986). *Test de Vocabulario Imágenes Peabody.* Circle Pines, MN: American Guidance Service.

Dziak, J. J., Lanza, S. T., & Tan, X. (2014). Effect size, statistical power, and sample size requirements for bootstrap likelihood ratio test in latent class analysis. *Structural*

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*(2), 190-193. https://doi.org/10.1177/1745691617720478

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*, 309-331. https://doi.org/10.1037/0096-3445.128.3.309

Etmanskie, J. M., Partanen, M., & Siegel, L. S. (2016). A longitudinal examination of the persistence of late-emerging reading disabilities. *Journal of Learning Disabilities, 49*(1), 21-35. doi:10.1177/0022219414522706

Farnia, F., & Geva, E. (2019). Late-emerging developmental language disorders in English-speaking monolinguals and English-language learners: A longitudinal perspective. *Journal of Learning Disabilities, 52*(6), 468-479. doi:10.1177/0022219419866645

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2010). Uncoupling of reading and IQ over time: Empirical evidence for a definition of dyslexia. *Psychological Science, 21*(1), 93-101. doi:10.1177/0956797609354084

Finders, J. K., McClelland, M. M., Geldhof, G. J., Rothwell, D. W., & Hatfield, B. E. (2021). Explaining achievement gaps in kindergarten and third grade: The role of self-regulation and executive function skills. *Early Childhood Research Quarterly, 54*, 72-85. doi:10.1016/j.ecresq.2020.07.008

Foster, M. E., Sevcik, R. A., Romski, M., & Morris, R. D. (2015). Effects of phonological awareness and naming speed on mathematics skills in children with mild intellectual disabilities. *Developmental Neurorehabilitation, 18*(5), 304-316. doi:10.3109/17518423.2013.843603

Francis, D. J., Rojas, R., Gusewski, S., Santi, K. L., Khalaf, S., Hiebert, L., & Bunta, F. (2019). Speaking and reading in two languages: On the identification of reading and language disabilities in Spanish-speaking English learners. *New Directions for Child and Adolescent Development, 2019*(166), 15-41. doi:10.1002/cad.20306

Friedman, N. P., Haber, B. C., Willcutt, E. G., Miyake, A. Young, S. Corely, R. P. & Hewitt, J. K. (2007). Greater attention problems during childhood predict poorer executive functioning in late adolescence. *Psychological Science, 18,* 893-900. https://doi.org/10.1111/j.1467-9280.2007.01997.x

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., Schatschneider, C., & Fletcher, J. M. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology, 98*, 29-43. doi:10.1037/0022-0663.98.1.29

Fuchs, L. S., Compton, D. L., Fuchs, D., Hollenbeck, K. N., Craddock, C. F., & Hamlett, C. L. (2008). Dynamic assessment of algebraic learning in predicting third graders' development of mathematical problem-solving. *Journal of Educational Psychology, 100*(4), 829-850. doi:10.1037/a0012657

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. N. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading, 22*(2), 152-166. doi:https://doi.org/10.1080/10888438.2017.1398259

Gaylord-Harden, N., Cunningham, J. A., Holmbeck, G. N., & Grant, K. E. (2010). Suppressor effects in coping research with African American adolescents from low-income communities. *Journal of Consulting and Clinical Psychology, 78*(6), 843-855. doi:10.1037/a0020063

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology, 104*(1), 206-223. doi:/10.1037/a0025398

Geary, D. C., Nicholas, A., Li, Y., & Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement:

An eight-year longitudinal study. *Journal of Educational Psychology, 109*(5), 680-693. doi:/10.1037/edu0000159

Getzmann, S., Wascher, E., & Schneider, D. (2018). The role of inhibition for working memory processes: ERP evidence from a short-term storage task. *Psychophysiology, 55*(5), 1-9. doi:10.1111/psyp.13026

Giofrè, D., Toffalini, E., Altoè, G., & Cornoldi, C. (2017). Intelligence measures as diagnostic tools for children with specific learning disabilities. *Intelligence, 61*, 140-145. doi:10.1016/j.intell.2017.01.014

Georgiou, G. K., Tziraki, N., Manolitsis, G., & Fella, A. (2013). Is rapid automatized naming related to reading and mathematics for the same reason(s)? A follow-up study from kindergarten to grade 1. *Journal of Experimental Child Psychology, 115*(3), 481-496. doi: 10.1016/j.jecp.2013.01.004

González, J. E. J., & Valle, I. H. (2000). Word identification and reading disorders in the Spanish language. *Journal of Learning Disabilities, 33*(1), 44-60. doi:10.1177/002221940003300108

Goodrich, J. M., & Namkung, J. M. (2019). Correlates of reading comprehension and word-problem solving skills of Spanish-speaking dual language learners. *Early Childhood Research Quarterly, 48*, 256-266. doi:10.1016/j.ecresq.2019.04.006

Gottardo, A., Collins, P., Baciu, I., & Gebotys, R. (2008). Predictors of grade 2 word reading and vocabulary learning from grade 1 variables in Spanish-speaking children: Similarities and differences. *Learning Disabilities Research & Practice, 23*(1), 11-24. doi:10.1111/j.1540-5826.2007.00259.x

Grimm, R.P., Solari, E.J., Gerber, M.M., Nylund-Gibson, K., & Swanson, H.L. (2019). A crosslinguistic examination of heterogeneous reading profiles of Spanish-speaking bilingual students. *The Elementary School Journal, 120,* 109-131.

Gunnerud, H. L., ten Braak, D., Reikerås, E.,Kirsti Lie, Donolato, E., & Melby-Lervåg, M. (2020). Is bilingualism related to a cognitive advantage in children? A systematic review and meta-analysis. *Psychological Bulletin, 146*(12), 1059-1083. doi:10.1037/bul0000301

Gorman, B. K. (2012). Relationships between vocabulary size, working memory, and phonological awareness in Spanish-speaking English language learners. *American Journal of Speech-Language Pathology, 21*(2), 109-123. https://doi.org/10.1044/1058-0360(2011/10-0063)

Harwell, M., & LeBeau, B. (2010). Student eligibility for a free lunch as an SES measure in education research. *Educational Researcher, 39*(2), 120-131. doi:10.3102/0013189X10362578

Hemphill, F.C., & Vanneman, A. (2011). Achievement Gaps: How Hispanic and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress (NCES 2011-459). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., & Trezise, K. (2018). Informative tools for characterizing individual differences in learning: Latent class, latent profile, and latent transition analysis. *Learning and Individual Differences, 66*, 4-15. doi:10.1016/j.lindif.2017.11.001

Hoff, E. (2018). Bilingual development in children of immigrant families. *Child Development Perspectives, 12*(2), 80-86. doi:10.1111/cdep.12262

Hosmer, D. and Lemeshow, S. (2000) Applied Logistic Regression (Second Edition). New York: John Wiley & Sons, Inc.

Hox, J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd Ed.) New York, NY: Routledge/Taylor & Francis.

Iglesias-Sarmiento, V., López, N. C., & Rodríguez, J. L. R. (2015). Updating executive function and performance in reading comprehension and problem solving. *Anales De Psicología, 31*(1), 298-309. doi:https://doi.org/10.6018/analesps.31.1.158111

Jordan, N. C., & Hanich, L. B. (2003). Characteristics of children with moderate mathematics deficiencies: A longitudinal perspective. *Learning Disabilities Research & Practice, 18*(4), 213-221. doi:10.1111/1540-5826.00076

Kaushanskaya, M., & Yoo, J. (2013). Phonological short-term and working memory in bilinguals' native and second language. *Applied Psycholinguistics, 34*(5), 1005-1037. doi:10.1017/S0142716412000100

Kam, C., Morin, A. J. S., Meyer, J. P., & Topolnytsky, L. (2016). Are commitment profiles stable and predictable? A latent transition analysis. *Journal of Management, 42*(6), 1462-1490. doi:10.1177/0149206313503010

Kaplan, D., & Keller, B. (2011). A note on cluster effects in latent class analysis. *Structural Equation Modeling, 18*(4), 525-536. doi:10.1080/10705511.2011.607071

Kempert, S., Saalbach, H., & Hardy, I. (2011). Cognitive benefits and costs of bilingualism in elementary school students: The case of mathematical word problems. *Journal of Educational Psychology, 103*(3), 547-561. https://doi.org/10.1037/a0023619

Kieffer, M. J., & Thompson, K. D. (2018). Hidden progress of multilingual students on NAEP. *Educational Researcher, 47*(6), 391-398. Retrieved from https://search.proquest.com/docview/2315229412?accountid=14521

Kim, Y. (2019). The causal structure of suppressor variables. *Journal of Educational and Behavioral Statistics, 44*(4), 367-389. doi:10.3102/1076998619825679

Lanza, S. T., Bray, B. C., & Collins, L. M. (2013). An introduction to latent class and latent transition analysis. In J. A. Schinka, W. F. Velicer & I. B. Weiner (Eds.), *2nd ed.; handbook of psychology: Research methods in psychology (vol. 2, 2nd ed.)* (2nd ed. ed., pp. 691-716, Chapter xxiv, 776 Pages) John Wiley & Sons, Inc., Hoboken, NJ. Retrieved from https://www.proquest.com/books/introduction-latent-class-transition-analysis/docview/1267038478/se-2?accountid=14521

Lanza, S. T., Dziak, J. J., Huang, L., Xu, S., & Collins, L. M. (2011). *Proc LCA and Proc LTA users' guide* (Version 1.2.7). University Park: The Methodology Center, Penn State. Retrieved from: http://methodology.psu.edu.

Lesaux, N. K., Lipka, O., & Siegel, L. S. (2006). Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading and Writing, 19*(1), 99-131. doi: 10.1007/s11145-005-4713-6

Lindholm-Leary, K. J. (2005). Review of research and best practices on effective features of dual language education programs. *Center for Applied Linguistics*, *35*, 17-23.

Lindholm-Leary, K., & Block, N. (2010). Achievement in predominantly low SES/Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism*, *13*(1), 43-60. https://doi.org/10.1080/13670050902777546

Maas, C., & Hox, J. (2005). Sufficient sample sizes in multiple regression analysis. *Methodology, 1,* 86-92. doi:10.1027/1614-2241.1.3.85

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism*, *11*(1), 81. doi:10.1017/S1366728907003227

Mann Koepke, K., & Miller, B. (2013). At the intersection of math and reading disabilities: Introduction to the special issue. *Journal of Learning Disabilities, 46*(6), 483-489. doi:10.1177/0022219413498200

Mancilla-Martinez, J., Hwang, J. K., Oh, M. H., & McClain, J. B. (2019). Early elementary grade dual language learners from Spanish-speaking homes struggling with English reading comprehension: The dormant role of language skills. *Journal of Educational Psychology,* doi10.1037/edu0000402

Martin, R. B., Cirino, P. T., Barnes, M. A., Ewing-Cobbs, L., Fuchs, L. S., Stuebing, K. K., & Fletcher, J. M. (2013). Prediction and stability of mathematics skill and difficulty. *Journal of Learning Disabilities, 46*(5), 428-443. https://doi.org/10.1177/0022219411436214

Masyn, K. (2013). Latent class analysis and finite mixture modeling. In T. Little (Ed.), *The Oxford handbook of quantitative methods in psychology* (Vol. 2, pp. 375-393). Oxford, UK: Oxford University Press.

McBride-Chang, C., Liu, P. D., Wong, T., Wong, A., & Shu, H. (2012). Specific reading difficulties in Chinese, English, or both: Longitudinal markers of phonological awareness, morphological awareness, and RAN in Hong Kong Chinese children. *Journal of Learning Disabilities, 45*(6), 503-514. doi: 10.1177/0022219411400748

McFatter, R. M. (1979). The use of structural equation models in interpreting regression equations including suppressor and enhancer variables. *Applied Psychological Measurement, 3*(1), 123-135. doi:10.1177/014662167900300113

Menon, V. (2016). Working memory in children's math learning and its disruption in dyscalculia. *Current Opinion in Behavioral Sciences, 10*, 125-132. https://doi.org/10.1016/j.cobeha.2016.05.014

Miciak, J., Ahmed, Y., Capin, P., & Francis, D. J. (2022). The reading profiles of late elementary english learners with and without risk for dyslexia. *Annals of Dyslexia,* doi:https://doi.org/10.1007/s11881-022-00254-4

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41(1),* 49-100. doi:10.1006/cogp.1999.0734

Morgan, P. L., & Farkas, G. (2016). Are we helping all the children that we are supposed to be helping? *Educational Researcher, 45*(3), 226-228. doi:/10.3102/0013189X16644607

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities, 42*(4), 306-321. doi: 10.1177/0022219408331037

Morgan, P. L., Farkas, G., Hillemeier, M. M., & Maczuga, S. (2017). Replicated evidence of racial and ethnic disparities in disability identification in U.S. schools. *Educational Researcher, 46*(6), 305-322. doi:10.3102/0013189X17726282

Morin, A. J. S., Bujacz, A., & Gagné, M. (2018). Person-centered methodologies in the organizational sciences: Introduction to the feature topic. *Organizational Research Methods, 21*(4), 803-813. doi:10.1177/1094428118773856

Muñoz-Sandoval, A. F., Woodcock, R. W., McGrew, K. S., & Mather, N. (2005). Bateria III Woodcock- Muñoz. Rolling Meadows, IL: Riverside.

Murphy, M.M., Mazzocco, M.M.M., Hanich, L.B., & Early, M.C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities*, 40(5), 458-478.

Muthén, B., & Asparouhov, T. (2020). Latent transition analysis with random intercepts (RI-LTA). *Psychological Methods,* doi:10.1037/met0000370

Muthén, B., & Muthén, L. K. (2012). Mplus:User's Guide. Los Angeles ,CA. Muthén & Muthén.

National Assessment of Educational Progress (2011). Achievement gap: How Hispanics and white students in public schools perform in mathematics and reading on the national assessment of educational progress. Washington DC: US Department of Education.

National Assessment of Educational Progress (2017). The condition of education (update 2017) Washington DC: US Department of Education.

National Assessment of Educational Progress (2019) NAEP Mathematics and Reading Assessments: Highlighted Results at Grades 4 and 8 for the Nation, States, and Districts Washington DC: US Department of Education.

National Center for Education Statistics (2020) Common Core of Data (CCD) "Local Education Agency Universe Survey. Washington DC: U.S. Department of Education

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd Ed.). NY: McGraw-Hill.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling, 14*(4), 535-569. doi:10.1080/10705510701575396

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science, 4*(4), 440-461. doi:10.1037/tps0000176

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. doi:10.1080/10705511.2014.915375

Purpura, D. J., Logan, J. A. R., Hassinger-Das, B., & Napoli, A. R. (2017). Why do early mathematics skills predict later reading? the role of mathematical language. *Developmental Psychology, 53*(9), 1633-1642. doi:https://doi.org/10.1037/dev0000375

Passolunghi, M. C., & Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology, 88*(4), 348-367. https://doi.org/10.1016/B978-0-12-410388-7.00010-5

Peña, E. D., Bedore, L. M., & Kester, E. S. (2016). Assessment of language impairment in bilingual children using semantic tasks: Two languages classify better than one. *International Journal of Language & Communication Disorders, 51*(2), 192-202. doi:10.1111/1460-6984.12199

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L.,Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin, 144*(1), 48-76. doi:/10.1037/bul0000124

Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016a). 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. doi:/10.1037/edu0000079

Peng, P., Namkung, J., Fuchs, D., Fuchs, L., Patton, S., Yen, L., Compton, D. L., Zhang, W. J., Miller, A. & Hamlett, C. (2016b). A longitudinal study on predictors of early calculation development among young children at risk for learning difficulties. *Journal of Experimental Child Psychology, 152*, 221-241. doi:/10.1016/j.jecp.2016.07.017

Prevoo, M. J. L., Malda, M., Mesman, J., & van IJzendoorn, M. H. (2016). Within- and cross-language relations between oral language proficiency and school outcomes in bilingual children with an immigrant background: A meta-analytical study. *Review of Educational Research, 86*(1), 237-276. doi:10.3102/0034654315584685

Quinn, J. M., Wagner, R. K., Petscher, Y., Roberts, G., Menzel, A. J., & Schatschneider, C. (2020). Differential codevelopment of vocabulary knowledge and reading comprehension for students with and without learning disabilities. *Journal of Educational Psychology, 112*(3), 608-627. doi:10.1037/edu0000382

Raven, J. C. (1976). *Colored Progressive Matrices*. London, England: H. K. Lewis & Co. Ltd.

Raudenbush, S.W., Yang, M.L., Yosef, M. (2000). Maximum likelihood for generalized linear models nested random effects via higher-order multivariate Laplace approximation. *Journal of Computational and Graphical Statistics, 9,* 141-157.

Rogosa, D. R., & Gottman, J. (Ed.) (1995). *The analysis of change*. Hillsdale, NJ: Erlbaum

Rojas, R., Hiebert, L., Gusewski, S., & Francis, D. J. (2019). Moving forward by looking back: Understanding why some Spanish-speaking English learners fall behind. *New Directions for Child and Adolescent Development, 2019*(166), 43-77. doi:10.1002/cad.20305

Ryoo J. H., Wang , C., Swearer, S. M., Hull, M., & Shi, D. (2018) Longitudinal Model Building Using Latent Transition Analysis: An Example Using School Bullying Data. Frontiers in Psychology, 9:675. doi: 10.3389/fpsyg.2018.00675

Sabourin, L., & Vīnerte, S. (2015). The bilingual advantage in the stroop task: Simultaneous vs. Early bilinguals. *Bilingualism: Language and Cognition, 18*(2), 350-355. doi:10.1017/S1366728914000704

Santi, K. L., Khalaf, S., Bunta, F., Rojas, R., & Francis, D. J. (2019). IQ-achievement discrepancy for identification of disabilities in Spanish-speaking English learners. *New Directions for Child and Adolescent Development, 2019*(166), 111-143. doi:10.1002/cad.20304

SAS Institute (2014). *SAS/STAT software: Changes and Enhancements through release 9.3*. Cary, NC: SAS Institute Inc.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development, 60*(4), 973-980. doi:10.2307/1131037

Spencer, M., Fuchs, L. S., Geary, D. C., & Fuchs, D. (2022). Connections between mathematics and reading development: Numerical cognition mediates relations between foundational competencies and later academic outcomes. *Journal of Educational Psychology, 114*(2), 273-288. doi:https://doi.org/10.1037/edu0000670

Snijders, T., & Bosker, R. (1999). *Multilevel modeling: An introduction to basic and advanced multilevel modeling.* Thousand Oaks, CA: Sage Press.

Snowling, M. J. (2012). Changing concepts of dyslexia: Nature, treatment and comorbidity. *Journal of Child Psychology and Psychiatry, 53*(9), e1-e3. doi:10.1111/j.1469-7610.2009.02197.x

Stanovich, K. E., & Siegel, L. (1994). Phenotypic performances profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Education Psychology, 86*(1), 24-53. doi:10.1037/0022-0663.86.1.24

Sullivan, A. L., Artiles, A. J., & Hernandez-Saca, D. (2015). Addressing special education inequity through systemic change: Contributions of ecologically based organizational consultation. *Journal of Educational & Psychological Consultation, 25*(2-3), 129-147. doi:10.1080/10474412.2014.929969

Swanson, H. L. (1992). Generality and modifiability of working memory among skilled and less skilled readers. *Journal of Educational Psychology, 84*(4), 473-488. doi:http://dx.doi.org/10.1037/0022-0663.84.4.473

Swanson, H.L. (2013). Abbreviated Test of Working Memory. American Psychological Association, Washington DC: PyscTESTS

Swanson, H. L. (2020). Specific learning disabilities as a working memory deficit: A model revisited. In A. J. Martin, R. A. Sperling & K. J. Newton (Eds.), *Handbook of educational psychology and students with special needs; handbook of educational psychology and students with special needs* (pp. 19-51, Chapter viii, 734 Pages). New York, NY: Routledge/Taylor & Francis Group.

Swanson, H. L., Arizmendi, G. D., & Li, J. (2021). The stability of learning disabilities among emergent bilingual children: A latent transition analysis. *Journal of Educational Psychology, 113*(6), 1244-1268. doi:http://dx.doi.org/10.1037/edu0000645

Swanson, H. L., & Fung, W. (2016). Working memory components and problem-solving accuracy: Are there multiple pathways? *Journal of Educational Psychology, 108*(8), 1153-1177. doi:10.1037/edu0000116

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 100*(2), 343-379. doi:10.1037/0022-0663.100.2.343

Swanson, H. L., Kudo, M., & Guzman-Orth, D. (2016). Cognition and literacy in English language learners at risk for reading disabilities: A latent transition analysis. *Journal of Educational Psychology, 108*(6), 830-856. doi:10.1037/edu0000102

Swanson, H. L., Kudo, M. F., & Van Horn, M. L. (2019). Does the structure of working memory in el children vary across age and two language systems? *Memory, 27*(2), 174-191. doi:10.1080/09658211.2018.1496264

Swanson, H. L., Kong, J. E., & Petcu, S. D. (2019). Individual differences in math problem solving and executive processing among emerging bilingual children. *Journal of Experimental Child Psychology, 187*, 25. doi:10.1016/j.jecp.2019.06.006

Swanson, H. L., Olide, A. F., & Kong, J. E. (2018). Latent class analysis of children with math difficulties and/or math learning disabilities: Are there cognitive differences? *Journal of Educational Psychology, 110*(7), 931-951. doi:10.1037/edu0000252

Swanson, H. L., Orosco, M. J., & Lussier, C. M. (2015). Growth in literacy, cognition, and working memory in English language learners. *Journal of Experimental Child Psychology, 132*, 155-188. doi:10.1016/j.jecp.2015.01.001

Swanson, H. L., Sáez, L., Gerber, M., & Leafstedt, J. (2004). Literacy and cognitive functioning in bilingual and nonbilingual children at or not at risk for reading disabilities. *Journal of Educational Psychology, 96*(1), 3-18. doi:10.1037/0022-0663.96.1.3

Swanson, H. L., Sáez, L., & Gerber, M. (2006). Growth in literacy and cognition in bilingual children at risk or not at risk for reading disabilities. *Journal of Educational Psychology, 98*(2), 247-264. doi:10.1037/0022-0663.98.2.247

Toll, S. W. M., Van, D. V., Kroesbergen, E. H., & Van Luit, Johannes E. H. (2011). Executive functions as predictors of math learning disabilities. *Journal of Learning Disabilities, 44*(6), 521-532. https://doi.org/10.1177/0022219410387302

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin, 109*(3), 524-536. doi:10.1037/0033-2909.109.3.524

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher-order abilities. *Psychological Bulletin, 133*(6), 1038-1066. https://doi.org/10.1037/0033-2909.133.6.1038

Vermunt, J. K. (2007). Growth models for categorical response variables: Standard, latent-class, and hybrid approaches. In K. van Montfort, J. Oud & A. Satorra (Eds.), *Longitudinal models in the behavioral and related sciences; longitudinal models in the behavioral and related sciences* (pp. 139-158) Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

Verhagen, J., & Leseman, P. (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. *Journal of Experimental Child Psychology, 141*, 65-82. doi:10.1016/j.jecp.2015.06.0156), 521-532.

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

Wagner, R., Torgesen, J., & Rashotte, C. (2000). *Comprehensive Test of Phonological Processes.* Austin TX: Pro-ED.

Wang, K., Banich, M. T., Reineberg, A. E., Leopold, D. R., Willcutt, E. G., Cutting, L. E., . . . Petrill, S. A. (2020). Left posterior prefrontal regions support domain-general executive processes needed for both reading and math. *Journal of Neuropsychology, 14*(3), 467-495. doi:10.111/jnp.12201

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-Third Edition*. San Antonio, TX: Psychological Corporation.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-Deficit/Hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 57*(11), 1336-1346. doi: 10.1016/j.biopsych.2005.02.006

Woodcock, R. W., McGrew, K. S., Schrank, F.A. , & Mather, N. (2007). Technical manual. Woodcock-Johnson III Normative Update. Rolling Meadows, IL: Riverside Publishing.

Woodcock, R. W., Muñoz-Sandoval, A. F. & Alvarado, C. G. (2005). *Woodcock-Muñoz Language Survey*. Itasca, IL: Riverside Publishing.

Yeung, S. S. (2018). Second language learners who are at-risk for reading disabilities: A growth mixture model study. *Research in Developmental Disabilities, 78*, 35-43. doi:10.1016/j.ridd.2018.05.001

Footnotes

[1]The new American Psychiatric Association (2013) *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) does not use the term "learning disabilities" or related terms (e.g., dyslexia, math disabilities, reading disabilities). The DSM-5 (2013) uses the term "specific learning disorder in reading" or "specific learning disorder in math" and assumes that such disorders have a neurological/biological base. In general, researchers use the terms such as specific reading disabilities and/or math disabilities to identify children at risk who are of average intelligence, but whose performance is below a certain percentile (e.g., 8th, 16th, 25th percentile) on a norm-referenced standardized reading and/or math measure (e.g., Cirino et al. 2015; Siegel & Stanovich, 1994). In terms of research, it is not uncommon to find children with normal intelligence defined as having a reading disability, but also experience serious difficulties in math and vice versa (e.g., Mann Koepke & Miller, 2013). This is because both reading and math draw upon or overlap with similar cognitive processes (Wang et al., 2020).

[2]A model to capture individual differences among latent classes is Baddeley's multicomponent working memory model (2012; Baddeley & Logie, 1999). According to Baddeley's model (Baddeley & Logie, 1999), WM is comprised of a central executive controlling system that interacts with a set of two subsidiary storage systems: the speech-based phonological loop and the visual-spatial sketchpad. The speech-based phonological loop is responsible for the temporary storage of verbal information; items are held within a phonological store of limited duration and are maintained within the store through a subvocal articulation process. The speech-based phonological loop is commonly associated with short-term memory (STM) because it involves two major components discussed in the STM literature: a speech-based phonological input store and a rehearsal process. The visual-spatial sketchpad is responsible for the storage of visual-spatial information over brief periods of time and plays a

key role in the generation and manipulation of mental images. The central executive is involved in the control and regulation of the WM system.

[3]These variables are called suppressors in regression modeling because they suppress the outcomes of irrelevant variance in other predictors causing the suppressed variables to obtain a substantial regression weight (e.g., Tzelgov & Henik, 1991). Prediction in the outcome variable (in this case latent class status) increases when a suppressor variable (in this case L1 language) is added to the equation simply because the suppressor variable is correlated with another predictor or set of predictors that are correlated with the outcome variable. In this case, the suppressor variable removes irrelevant predictive variance from other variables and increases the predictor's weight thus increasing overall model predictability.

Table 1

Variation in Normative Scores for Manifest Variables from Wave 1 to Testing Wave 3

	Total Sample			Cohort 1			Cohort 2			Cohort 3		
Variable	N	M	SD	N	M	SD	N	M	SD	N	M	SD
Vocabulary												
E-PPVT	266	-1.00	15.80	109	2.66	15.68	81	-0.95	16.97	76	**-6.29**	13.16
E-expressive	263	4.57	21.86	108	4.30	24.79	80	4.75	20.03	75	4.77	19.38
S-TVIP	255	-0.87	18.74	107	-2.95	15.93	78	**-5.81**	19.54	70	7.80	19.16
S-expressive	263	0.73	14.11	107	0.30	16.32	80	-0.09	12.5	76	2.18	12.29
Math												
E-problem-solving	257	**-5.71**	16.20	109	**-12.55**	14.56	79	-1.34	14.74	69	0.09	16.44
E-calculation	262	**-5.91**	15.21	106	**-8.36**	18.69	82	**-5.67**	10.34	74	-2.69	13.66
S-problem-solving	247	**-7.11**	15.59	107	**-11.77**	14.03	75	-4.72	16.41	65	-2.20	15.16
S-calculation	257	**-13.64**	14.50	105	**-19.86**	11.85	80	**-10.73**	13.09	72	**-7.79**	16.13
Reading												
E-wordID	255	1.21	12.66	106	-0.12	13.93	77	1.87	13.46	72	2.48	9.36
E-compre	258	-3.90	15.93	108	**-5.23**	18.91	80	**-5.15**	14.08	70	-0.43	12.10
S-wordID	227	-3.52	16.65	96	**-7.88**	15.84	73	-1.73	15.78	58	1.44	17.48
S-compre	248	**-13.52**	14.15	104	**-16.00**	15.32	75	**-13.37**	12.12	69	**-9.96**	13.79
Attention/Fluid Intelligence												
CTRS	198	0.01	9.14	86	0.88	9.74	74	-0.92	8.45	38	-0.20	9.07
RAVEN	247	1.05	18.36	100	3.05	18.79	83	0.88	19.23	64	-1.85	16.29

Cohort 1=grades 1-3, Cohort 2=grades 2-4, Cohort 3-grades 3-5; E-English, S-=Spanish, PPVT=Peabody Picture Vocabulary Test, TVIP= The Test de Vocabulario en Imágenes Peabody; Expressive=One-Word Expressive Vocabulary Test, WPS=Word Problem Solving test from WJ or Batería, Calculation=Arithmetic Subtest from WJ or Batería, WordID=Word Identification subtest from Woodcock-Muñoz, Compre=Passage Comprehension subtest from Woodcock-Muñoz. CTRS=Conners Behavior Rating Scale (T-score), Fluid Int.=Raven Colored Progressive Matrices Test. **Bold**=A differences of 5 or more standard scores from Wave 1 to Wave 3.

Table 2

Fit Indices for Seven Latent Class Models for Year 1 (Wave 1) Year 2 (Wave 2). And Year 3(Wave 3)

	Wave	LC=1	LC=2	LC=3	LC=4	LC=5	LC=6	LC=7
Year 1 (Wave 1)								
	Log-likelihood:	-2096.97	-1999.35	-1924.81	-1901.88	-1890.74	-1868.04	-1862.44
	G-squared:	1368.05	1172.82	1023.74	977.88	955.6	910.2	898.99
	AIC:	1396.05	1230.82	1111.74	**1095.88**	1103.6	1088.2	1106.99
	BIC:	1447.14	1336.64	1272.3	**1311.17**	1373.62	1412.96	1486.48
	CAIC:	1461.14	1365.64	1316.3	**1370.17**	1447.62	1501.96	1590.48
	Adjusted BIC:	1402.75	1244.68	1132.77	**1124.08**	1138.97	1130.74	1156.69
	Entropy:	1	0.79	0.8	0.85	0.82	0.9	0.81
	Degrees of freedom:	16369	16354	16339	16324	16309	16294	16279
	LMR	0	0	0	0.034	0.45	0.57	0.76
	Bootstrap	0	0		0	1	1	1
Year 2 (Wave 2)								
	Log-likelihood:	-2189.94	-2017.35	-1908.6	-1865.01	-1844.49	-1828.71	-1807.49
	G-squared:	1568.1	1222.92	1005.43	918.25	877.21	845.64	803.19
	AIC:	1596.1	1280.92	1093.43	1036.25	1025.21	1023.64	1011.19
	BIC:	1647.19	1386.74	1253.99	**1251.54**	1295.23	1348.4	1390.69
	CAIC:	1661.19	1415.74	1297.99	**1310.54**	1369.23	1437.4	1494.69
	Adjusted BIC:	1602.79	1294.78	1114.46	1064.45	1060.58	1066.18	1060.9
	Entropy:	1	0.85	0.84	0.84	0.83	0.88	0.88
	Degrees of freedom:	16369	16354	16339	16324	16309	16294	16279
	LMR (p values)	0			0.006	0.03	0.77	0.76
	Bootstrap	0			0	0	1	0.5
Year 3 (Wave 3)								
	Log-likelihood:	-2234.66	-2062.19	-1988.62	-1958.83	-1923.02	-1909.08	-1893.25
	G-squared:	1616.11	1271.17	1124.04	1064.45	992.83	964.96	933.3
	AIC:	1644.11	1329.17	1212.04	1182.45	1140.83	1142.96	1141.3

	BIC:			1694.69	1433.95	1371.01	**1395.62**	1408.2	1464.53	1517.07
	CAIC:			1708.69	1462.95	1415.01	**1454.62**	1482.2	1553.53	1621.07
	Adjusted BIC:			1650.3	1342	1231.5	1208.55	1173.56	1182.33	1187.31
	Entropy:			1	0.88	0.85	0.87	0.81	0.82	0.82
	Degrees of freedom:			16369	16354	16339	16324	16309	16294	16279
	LMR			0	0	0	0	0.32	0.056	0.285
	Bootstrap			0	0	0	0	0.16	0	1

Note. LC=Latent Class, AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; CAIC and Adjusted BIC corrected for sample size; LMR = Lo-Mendell-Rubin Test; BLRT = Bootstrap Likelihood Ratio Test. **Bold**= elbow in index

Table 3
Delta Estimates Status Membership Probabilities

Total Sample				
Latent Class	**1**	**2**	**3**	**4**
Year 1 :	0.04	0.45	0.2	0.31
Year 2 :	0.10	0.22	0.22	0.47
Year 3 :	0.19	0.11	0.23	0.47
Cohort 1				
Latent Class	1	2	3	4
Year 1 :	0.02	0.67	0.13	0.18
Year 2 :	0.08	0.42	0.18	0.32
Year 3 :	0.16	0.17	0.25	0.43
Cohort 2				
Status:	1	2	3	4
Year 1 :	0.05	0.39	0.27	0.29
Year 2 :	0.14	0.10	0.23	0.52
Year 3 :	0.28	0.06	0.22	0.44
Cohort 3				
Status:	1	2	3	4
Year 1 :	0.07	0.27	0.19	0.48
Year 2 :	0.06	0.13	0.23	0.58
Year 3 :	0.11	0.10	0.22	0.57

LC1=poor achiever-comorbid LD, LC2= balanced bilingual average achiever, LC3=children at risk for math problem-solving disabilities, LC4=English dominant Average achiever,Cohort 1= grades 1 to 3; Cohort 2=grades 2 to 4; Cohort 3=grades 3 to 5

Table 4

Transition Probabilities by total Sample and Cohort

	Year 1(rows) to Year 2 (Columns)				Year 2 (rows) to Year 3 (Columns)			
Total Sample								
Latent Class	1	2	3	4	1	2	3	4
1	1.00	0	0	0	0.86	0	0	0.14
2	0	0.49	0.13	0.38	0	0.51	0.14	0.36
3	0.22	0	0.78	0	0.07	0	0.93	0
4	0.05	0	0	0.95	0.19	0	0	0.81
Cohort 1								
Latent Class	1	2	3	4	1	2	3	4
1	1.00	0	0	0	0.88	0	0	0.12
2	0	0.62	0.13	0.25	0.02	0.39	0.20	0.39
3	0.27	0	0.73	0	0.11	0	0.89	0
4	0.13	0.05	0	0.81	0.19	0	0	0.81
Cohort 2								
Latent Class	1	2	3	4	1	2	3	4
1	1.00	0	0	0	0.85	0	0	0.15
2	0	0.27	0.14	0.59	0	0.56	0	0.44
3	0.35	0	0.65	0	0.04	0	0.95	0.02
4	0.01	0	0	0.99	0.29	0	0	0.71
Cohort 3								
Latent Class	1	2	3	4	1	2	3	4
1	0.86	0	0.14	0	1.00	0	0	0
2	0	0.50	0.10	0.40	0	0.70	0.05	0.25
3	0	0	1.00	0	0.06	0	0.94	0
4	0	0	0	1.00	0.07	0.01	0	0.91

LC1=poor achiever or comorbid LD, LC2= balanced bilingual average achiever, LC3=children at risk for math problem-solving disabilities, LC4=English dominant-average achiever,

Cohort 1= grades 1 to 3; Cohort 2=grades 2 to 4; Cohort 3=grades 3 to 5

Table 5

Means and Standard Deviations of Manifest Variables as a Function of Latent Class at Wave 3 and Difference Scores (Wave 3 minus Wave 1)

Latent Class		LC1			LC2			LC3			LC4	
Variable	N	M	SD	N	M	SD	N	M	SD	N	M	SD
Age (mos)	24	112.33	10.32	69	111.18	13.50	58	115.32	11.94	118	119.67	13.87
E-ppvt	24	**83.14**	17.24	68	100.27	16.37	56	**77.71**	12.07	118	99.89	12.07
E-expressive	24	94.59	21.47	67	117.00	21.87	56	**88.52**	16.50	118	120.09	18.83
S-ppvt	22	**82.29**	13.29	65	103.02	16.48	57	97.07	15.42	112	90.67	17.13
S-expressive	24	**74.59**	10.9	66	89.56	18.29	56	92.26	10.81	118	70.41	10.44
E-problemsolving	24	**73.20**	9.85	64	99.21	14.10	56	**83.79**	14.85	113	96.27	14.16
E-calculation	24	94.01	14.13	67	114.92	10.20	55	104.84	11.31	118	111.86	11.14
S-problemsolving	22	**75.20**	13.54	62	106.78	11.16	57	93.16	10.61	107	97.14	12.33
S-calculation	23	**65.08**	12.7	67	91.10	11.5	54	**78.21**	14.53	118	87.76	14.32
E-wordID	24	**81.30**	15.58	63	113.76	10.93	55	100.42	11.85	114	108.26	12.38
E-compre	24	**69.20**	13.98	68	99.33	13.21	56	**80.22**	13.93	117	95.11	12.03
S-wordID	22	92.12	25.05	52	121.70	9.76	53	118.16	10.14	102	112.38	12.81
S-compre	21	**66.31**	11.97	65	90.49	9.55	55	**87.91**	8.46	111	72.18	12.86
CTRS	18	51.88	10.06	54	47.89	8.00	40	51.48	9.41	98	50.76	9.85
RAVEN	24	88.28	15.59	66	105.76	12.79	54	93.83	14.19	116	105.29	12.52
						Difference						
Variable	N	M	SD	N	M	SD	N	M	SD	N	M	SD
E-ppvt	24	0.96	19.17	68	1.48	17.41	56	-1.82	14.62	118	-2.43	14.58
E-expressive	24	2.01	19.39	66	7.34	26.57	56	10.42	15.64	117	0.74	21.37
S-ppvt	22	-4.73	18.22	65	-0.98	16.62	56	-4.32	16.64	112	1.67	20.72
S-expressive	24	3.18	14.94	66	-1.72	17.66	55	0.27	12.94	118	1.81	12.06
E-problemsolving	24	**-13.33**	11.21	64	-6.43	14.16	56	**-8.95**	15.95	113	-2.08	17.46
E-calculation	23	**-7.13**	16.7	66	-5.23	17.28	55	**-6.23**	18.75	118	-5.91	11.61
S-problemsolving	22	**-17.57**	15.99	62	-6.16	14.02	57	**-9.65**	13	106	-4.13	16.68
S-calculation	22	**-19.4**	19.82	66	-14.86	11.93	54	**-17.9**	12.49	115	-9.84	14.69
E-wordID	24	-1.77	16.26	63	0.52	10.67	55	2.11	13.76	113	1.8	12.33

E-comprehension	24	-3.05	15.57	66	-6.01	14.5	53	-4.26	16.72	115	-2.70	16.47
S-wordID	22	**-8.46**	21.87	52	-10.01	14.72	51	-3.76	17.54	102	0.97	14.56
S-comprehension	21	**-12.5**	14.51	65	-14.57	14.31	51	**-10.78**	10.6	111	-14.36	15.37
CTRS	15	-2.02	11.43	54	0.61	7.36	37	-0.16	6.37	92	0.03	10.59
Raven	21	3.23	23.36	65	0.72	19.82	49	-1.02	17.37	112	1.74	16.98

Note. Manifest variables are in Normed Scores (M=100, *SD*=15). E=English, S=Spanish. LC1=poor achiever or comorbid LD, LC2= balanced bilingual average achiever, LC3=children at risk for math problem-solving disabilities, LC4=English dominant-average achiever, PPVT=Picture Vocabulary Test, Expressive=One-Word Expressive Vocabulary Test, Problem solving=Word Problem-Solving test from WJ or Batería, Calculation=Arithmetic Subtest from WJ or Batería, Word=Word Identification subtest from Woodcock-Muñoz, Comprehension=Passage comprehension subtest from Woodcock-Muñoz. CRTS=Conners Behavior Rating Scale (T-score), Raven.=Raven Colored Progressive Matrices Test

Difference= Score at Wave 3 minus Score at Wave 1. **Bold**=means scores of risk groups (< 85 standard score) or negative difference scores (wave 3 – wave 1) > -5.00.

Table 6. Effect Sizes on Manifest Variables as a Function of Latent Class at Wave 3

Measure	LC1 vs. LC2	LC1 vs. LC3	LC1 vs. LC4	LC2 vs. LC3	LC2 vs. LC4	LC3 vs. LC4
Age (mos)	0.04	-0.32	-0.61	-0.33	-0.62	-0.32
E-ppvt	-1.03[a]	0.39	**-1.28**	1.55	0.03	**-1.84**
E-express	-1.03	0.34	**-1.32**	1.45	-0.15	**-1.74**
S-ppvt	-1.32	-0.99	-0.51	0.37	**0.73**	0.39
S-express	-0.90	-1.63	0.40	-0.18	**1.39**	**2.07**
E-problem	-1.99	-0.78	**-1.71**	1.07	0.21	**-0.87**
E-calculation	-1.84	-0.89	**-1.53**	0.94	0.28	-0.63
S-problem	-2.67	-1.56	**-1.75**	1.25	**0.81**	-0.34
S-calculation	-2.20	-0.94	**-1.61**	1.00	0.25	-0.66
E-wordID	-2.63	-1.46	**-2.08**	1.17	0.46	-0.64
E-compre	-2.25	-0.79	**-2.09**	1.41	0.34	**-1.18**
S-wordID	-1.87	-1.63	**-1.30**	0.36	**0.78**	0.48
S-compre	-2.38	-2.26	-0.46	0.28	**1.56**	**1.36**
CTRS	0.47	0.04	0.11	-0.42	-0.31	0.07
RAVEN	-1.29	-0.38	-1.30	0.89	0.04	**-0.88**

Note. [a]a negative effect size under LC1-LC2 means LC1 has a lower performance than LC2; LC1=poor achiever or comorbid LD, LC2= balanced bilingual average achiever, LC3=children at risk for math problem-solving/reading comprehension disabilities, LC4=English dominant-average achiever, Bold=Effect size .50 or greater for the learning disability latent classes. E-English, S-=Spanish, PPVT=Picture Vocabulary Test, Express=One-Word Expressive Vocabulary Test, Problem=Word Problem-Solving test from WJ or Batería, Calculation=Arithmetic Subtest from WJ or Batería, WordID=Word Identification subtest from Woodcock-Muñoz, Compre=Passage comprehension subtest from Woodcock-Muñoz. CTRS=Conners Behavior Rating Scale (T-score), Raven.=Raven Colored Progressive Matrices Test. **Bold**=effect sizes > .80 comparing risk groups (LC1, LC3) to Reference status group (LC4)

Table 7

Latent Class as a Function of Item Response Probabilities (Rho Estimates)

Latent Class	LC1	LC2	LC3	LC4
Language				
E-PPVT	0.77	0.36	0.88	0.17
E-expressive	0.57	0.19	0.60	0.07
S-TVIP	0.82	0.17	0.26	0.59
S-expressive	0.77	0.31	0.15	0.86
Math				
E-WPS	0.91	0.07	0.61	0.25
E-calculation	0.41	0.03	0.11	0.02
S-WPS	0.88	0.02	0.26	0.29
S-calculation	0.92	0.15	0.54	0.44
Reading				
E-wordID	0.79	0.04	0.22	0.07
E-compre	1.00	0.15	0.75	0.20
S-wordID	0.54	0.03	0.06	0.12
S-compre	1.00	0.07	0.45	0.89
Fluid Intell./Attention				
Raven	0.61	0.22	0.38	0.13
CTRS	0.21	0.09	0.20	0.17

E-English, S-Spanish, LC1=poor achiever or comorbid LD, LC2= balanced bilingual average achiever, LC3=children at risk for math problem-solving/reading comprehension disabilities, LC4=English dominant-average achiever, PPVT=Peabody Picture Vocabulary Test, TVIP= The Test de Vocabulario en Imágenes Peabody, WPS=Word Problem Solving test from WJ or Batería, Calculation=Arithmetic Subtest from WJ or Batería, WordID=Word Identification subtest from Woodcock-Muñoz, Compre=Passage Comprehension subtest from Woodcock-Muñoz. CTRS=Conners Teacher Behavior Rating Scale, Fluid Int.=Raven Colored Progressive Matrices Test.

Table 8
Generalized Linear Polytomous Model Predicting Latent Class at Wave 3 from Intercept and Slopes of Cognitive Measures

		LC1 vs. LC4			LC2 vs. LC4			LC3 vs. LC4		
				Empty Model						
Fixed Effects		Estimate	SE		Estimate	SE		Estimate	SE	
	Intercept	-1.96***	0.46		-0.46	0.40		-1.46**	0.58	
Random Effect		0.32	0.18		0.40	0.22		0.67	0.35	
				English Model						
Fixed Effects										
	Intercept	-5.05***	1.35		-0.45	0.46		-2.28***	0.71	
	E-STM	-0.47	0.37		-0.02	0.10		-0.33*	0.17	
	E-Speed	1.36**	0.41		-0.06	0.23		0.44	0.32	
	E-Inhib	-0.60*	0.29		0.15	0.10		-0.01	0.19	
	E-WM	-0.85***	0.30		0.16	0.11		-0.68***	0.14	
	VISWM	-0.06	0.18		0.05	0.10		-0.16	0.11	
Growth										
	E-STM	0.12	0.16		0.02	0.08		0.10	0.12	
	E-Speed	0.12	0.21		-0.03	0.13		0.01	0.17	
	E-Inhib	-0.21	0.16		0.05	0.09		-0.02	0.1	
	E-WM	0.28	0.28		0.13	0.13		0.16	0.14	
	VISWM	0.03	0.19		-0.09	0.11		-0.31*	0.14	
Random Effect-Level 2										
		0.84	0.61		0.54*	0.27		0.77**	0.31	
				Spanish Model						
Fixed Effects										
	Intercept	-4.74***	0.78		-0.41	0.45		-1.25*	0.54	
	S-STM	-0.48***	0.16		0.09	0.10		-0.04	0.15	

	S-Speed	0.23	0.33		-0.57*	0.25		-0.54	0.29	
	S-Inhib	-0.90	0.63		0.14	0.21		-0.11	0.19	
	S-WM	-0.67**	0.27		0.53**	0.12		-0.09	0.12	
	VISWM	-0.41*	0.17		-0.04	0.09		-0.15	0.11	
Growth										
	S-STM	0.35*	0.14		-0.06	0.09		0.07	0.09	
	S-Speed	-0.11	0.17		-0.07	0.09		-0.07	0.15	
	S-Inhib	-0.33	0.45		0.25	0.19		0.25	0.23	
	S-WM	-0.11	0.19		0.11	0.08		-0.13	0.10	
	VISWM	0.06	0.17		-0.06	0.09		-0.13	0.10	
Random Effect-Level 2										
		0.61	0.36		0.48**	0.19		0.49	0.27	

LC1=poor achiever or comorbid LD, LC2= balanced bilingual average achiever, LC3=children at risk for math problem-solving/reading comprehension disabilities, LC4=English dominant-average achiever,
E-English, S-=Spanish, STM=Short-Term Memory or Phonological Loop; Speed=Naming Speed, Inhib=Inhibition or Random Generation Tasks, WM=Executive component of working memory, Viswm=Visual Spatial working memory.
Empty model fit indicates for -2LL, AIC and BIC were 1762.21,1774.21, 1781.02, for English only model the fit indices were 1153.55, 1225.55 and 1266.42, and for the Spanish model they were 1184.73,1256.73.1296.01 respectively.
**p < .05, **p < .01, ***p < .001*

Table 9

Generalized Linear Polytomous Model Predicting Latent Class at Wave 3

		LC1 vs. LC4			LC2 vs. LC4			LC3 vs. LC4		
Fixed Effects		Estimate	SE		Estimate	SE		Estimate	SE	
	Intercept	-5.97***	1.32		-0.61	0.52		-1.94***	0.62	
	E-STM	-1.00***	0.34		-0.26*	0.12		-0.45**	0.16	
	S-STM	-0.09	0.25		0.14	0.12		0.29	0.16	
	E-Speed	1.03	0.54		0.46**	0.17		1.15***	0.27	
	S-speed	-0.09	0.59		-0.87***	0.25		-1.65***	0.40	
	E-Inhib	-0.59	0.59		-0.03	0.17		-0.03	0.16	
	S-Inhib	-0.25	0.69		0.08	0.27		0.04	0.20	
	E-WM	-0.80**	0.33		0.17	0.12		-0.73***	0.20	
	S-WM	-0.28	0.43		0.59***	0.14		0.01	0.14	
	VISWM	-0.21	0.20		-0.04	0.12		-0.19	0.11	
Growth										
	E-STM	0.01	0.35		-0.08	0.10		-0.01	0.14	
	S-STM	0.35*	0.15		-0.12	0.10		0.18	0.10	
	E-Speed	0.12	0.39		0.14	0.13		0.13	0.18	
	S-speed	-0.29	0.48		-0.12	0.12		-0.34	0.18	
	E-Inhib	0.09	0.30		-0.11	0.15		0.01	0.13	
	S-Inhib	-0.33	0.51		0.44	0.28		0.33	0.30	
	E-WM	0.90*	0.38		0.33***	0.11		0.33	0.18	
	S-WM	-0.55	0.32		0.04	0.11		-0.37**	0.14	
	VISWM	-0.20	0.24		-0.09	0.12		-0.25	0.15	
Random Effect-Level 2										
		0.56	0.34		0.32*	0.13		0.33*	0.16	

LC1=poor achiever or comorbid LD, LC2= balanced bilingual average achiever, LC3=children at risk for math problem-solving/reading comprehension disabilities, LC4=English dominant-average achiever,
E-English, S-=Spanish, STM=Short-Term Memory or Phonological Loop; Speed=Naming Speed, Inhib=Inhibition or Random Generation Tasks, WM=Executive component of working memory, Viswm=Visual Spatial working memory. The fit indices were 8336.30, 956.3 and 1021.76 for -2LL, BIC and AIC, respectively. **p* < .05, ***p* < .01, ****p* < .001