# Speaking and reading in two languages: On the identification of reading and language disabilities in Spanish-speaking English learners

**David J. Francis**

Department of Psychology, and Texas Institute for Measurement, Evaluation, Statistics, University of Houston, 4849 Calhoun Road, Houston, TX 77204-6022

**Raúl Rojas**

Communication Sciences and Disorders Program, School of Behavioral Sciences, University of Texas at Dallas, 1966 Inwood Rd., Dallas, TX 75235

**Svenja Gusewski**

Communication Sciences and Disorders Program, School of Behavioral Sciences, University of Texas at Dallas, 1966 Inwood Rd., Dallas, TX 75235

**Kristi L. Santi**

Educational Leadership and Policy Studies, University of Houston,

**Shiva Khalaf**

Department of Psychological, Health and Learning Sciences, University of Houston, 3657 Cullen Blvd., Houston, TX 77204-5029

**Lindsey Hiebert**

Communication Sciences and Disorders Program, School of Behavioral Sciences, University of Texas at Dallas, 1966 Inwood Rd., Dallas, TX 75235

**Ferenc Bunta**

Department of Communication Sciences and Disorders, University of Houston, 4455 Cullen Blvd, Houston, TX 77204-6018

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. In D. J. Francis (Ed.), *Identification, Classification, and Treatment of Reading and Language Disabilities in Spanish-speaking EL Students. New Directions for Child and Adolescent Development*, *166*, 15–41.

This work was supported by grant R324A160258 funded by the National Center for Special Education Research in the Institute of Education Sciences, and P50 HD052117, Texas Center for Learning Disabilities, from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The data reported here were collected under cooperative agreement R305P030031 funded by the US Department of Education and grant P01 HD39521 funded by NICHD. The attitudes and opinions expressed herein are those of the authors and do not reflect the position of the funding agencies or the Federal Government of the United States. We would like to thank the many teachers, students, parents, and colleagues who contributed to the original projects that made this research reported in this volume possible.

**Abstract**

Articles in this issue examine (1) the primary sources of variability in reading and language achievement among Spanish-speaking English learners (ELs) in the United States, (2) the extent to which poor performance at the end of grade 2 is identifiable in developmental trajectories beginning in kindergarten, (3) the relations among core reading constructs of phonological awareness and decoding in both English and Spanish and the factors that affect their relationship, (4) the performance of different approaches to identification and the factors that influence how well they work, as well as (5) the growing literature focused on intervention for reading problems in this population. This article examines the literature on language minority students and disability identification and analyzes a large-scale longitudinal dataset (>4,000 ELs; > 15,000 observations) to systematically characterize and describe the oral language and reading development of Spanish-speaking children designated as ELs from kindergarten to second grade, considering a range of factors that may potentially contribute to that characterization and its relation to academic performance. This systematic characterization should facilitate the development of an empirical basis for a theoretically grounded framework of typical development in ELs in order to more precisely identify those children with language and learning disabilities.

# Speaking and reading in two languages: Variance in language and literacy among Spanish-speaking English learners

Children growing up acquiring two languages differ in various ways from their monolingual peers, including the ways in which they learn to read. For example, in a large-scale, cross-sectional study of bilingual (Spanish-English) students in kindergarten to grade 3, Miller et al. (2006) found that oral language skills in each language predicted reading scores (reading comprehension and word reading efficiency) within and across languages. Similarly, studies including bilingual children have found cross-linguistic transfer regarding phonological awareness skills that, in turn, predict reading in both languages (Geva & Wang, 2001). These findings may be explained in part by the Branum-Martin et al. (2012) meta-analysis of the correlation between phonological awareness measured in English and in other languages, which estimated mean correlations ranging from .82 to .88 depending on the grain size of the language. Importantly, these are estimated mean correlations at the task level, which underestimate the correlation between constructs due to task unreliability.

Current guidelines for identifying children with reading and language disabilities in the US have been developed using data largely from monolingual English-speaking children. These guidelines may need to be adjusted for children who are acquiring more than one language (e.g., Spanish-English bilingual children), if for no other reason than that these children experience a variety of instructional models that are unique to them, such as bilingual education models of various forms, and structured immersion, where the instruction is delivered in a language in which the child is not yet proficient. Consequently, there is a critical need to better understand the oral language and literacy development of ELs in order to develop a theoretically grounded framework for defining reading and language disabilities and to empirically determine those

factors that impact the processes of identification, classification, and treatment in this educationally at-risk population -- issues that the present study addresses.

Students residing in the United States (US), but speaking a language other than English, comprise the fastest growing subgroup of students in the US (McFarland, et al., 2017). These students are collectively referred to as language minority students, and those who enter school not yet proficient in English are referred to as English Learners (ELs). Presently, US federal law allows for the identification of reading and language disabilities in students who are ELs using the same approaches that are accepted for native English speakers without explicit guidance on successful implementation of these models with EL students.

**Language Minority Students, English Learners, and Their Academic Achievement**

Numerous studies in the US have shown that on average, children who are designated as ELs lag behind their non-EL peers in academic achievement (Hoff, 2013; National Center for Education Statistics, 2015; Oller & Eilers, 2002). Although the number of languages spoken by language minority children in US public schools is estimated to exceed 400, approximately 77% have Spanish as their first language (McFarland, et al., 2017) with the next-most prevalent language being Arabic (2.3%). Consequently, Spanish-speaking ELs are of particular interest to educators and researchers in the US (Basterra, Trumbull, & Solano-Flores, 2010; Swanson, 2009). The reading skills of nearly 70% of Spanish-speaking ELs in fourth grade are below basic levels (as opposed to 27% of non-ELs), negatively impacting overall academic success for these children. Importantly, this longstanding gap in reading skills between EL and non-EL students still persists in 12$^{th}$ grade (National Center for Education Statistics, 2015).

This difference in achievement between EL and non-EL students is well documented, but is also likely biased against ELs for several reasons. Foremost among these reasons is an

accounting issue. Most research on the gap compares children currently designated as ELs to children not currently designated as EL. On the surface, this practice seems reasonable, but upon deeper inspection is found to be seriously flawed because students who began school as ELs and gained proficiency in English are omitted from the EL category; the practice thus essentially compares students not proficient in the societal language to children who are proficient in the societal language. Because language proficiency relates to achievement even among native speakers of a language, it is not surprising that such a comparison shows a gap in achievement for ELs. The magnitude of this bias is greater in later grades and has been termed "the Gap that Can't Go Away" (Saunders & Marcelletti, 2013). Indeed, when the achievement of children who were ever EL is compared to the achievement of children who were never EL, the achievement gap is typically much smaller or non-existent (Umansky, Thompson, & Diaz, 2017). At the same time, other factors also contribute, such as differences in economic advantage, with ELs more likely to live in poverty and not to have at least one parent who had graduated high school (McFarland et al., 2017).

The use of a single label to describe this group of students belies their heterogeneity. A growing body of research (Anthony et al., 2009; Hammer, Lawrence, & Miccio, 2007; Oller & Eilers, 2002; Páez, Tabors, & López, 2007) shows a variety of factors are associated with the developmental and academic outcomes of ELs. These include age of acquisition of each language, simultaneous or sequential bilingualism, the generation of immigration of the child and parent(s), the degree and quality of input in each language, socioeconomic status, parental education level, language of instruction in academic programs, interruption of participation in formal educational programs, among others (Hammer et al., 2014). Clearly, this heterogeneity makes the study of the language and literacy skills of ELs a complex enterprise, one in which

English outcomes often take precedence over considering the child's abilities in both languages. Although academic success in English is the end goal for children in the US, it is critical to consider the role of the native language in any investigation of ELs, as it is co-developing along with their acquisition of English, albeit differently depending on the home and community language context, and the language of instruction model(s) under which the child has been instructed.

All developed countries experience some degree of linguistic diversity among their populations and there is a sense that the phenomenon has increased across the globe in the 21st century. The percentage of foreign-born students participating in PISA grew from 9.4% to 12.5% between 2006 and 2015, with growth in both first-generation (4.5% to 5.4%) and second-generation (5.0% to 7.1%) immigrant students. Among first-generation immigrants, 67% spoke a language other than the language of the assessment at home, whereas among second-generation immigrants, 44.7% spoke a language other than the language of the assessment at home. At the same time, countries vary in the percentage of foreign-born and language minority students in their school systems, and in their success in equalizing educational outcomes for these students (OECD 2016). Both in the US and globally, language minority students and students who are first- or second-generation immigrants tend to be disadvantaged socio-economically and tend to score more poorly on these international assessments of Reading, Math, and Science.

Regardless of the demographic, social, and accounting practices that contribute to the achievement gap for language minority students in general, and ELs in particular, a critical step in closing the achievement gap is to better understand what characterizes students' oral language and reading skills in both languages during the early school years, because early language and literacy skills are highly predictive of later reading abilities and academic achievement (Halle,

Hair, Wandener, McNamara, & Chien, 2012; Han, 2012, Miller et al., 2006). To be comprehensive, this characterization must take into account the variance in oral language and reading skills within English and the native language. Previous work examining the oral language skills of Spanish-English bilinguals at 4 and 5 years of age indicate that they tend to score one-to-two standard deviations below monolingual norms in English receptive and expressive vocabulary and in auditory comprehension (Hammer, Lawrence, & Miccio, 2007; Páez, Tabors, & López, 2007). In emergent literacy skills such as phonological awareness and letter identification, Spanish-speaking bilingual preschoolers and kindergarteners score lower than their monolingual English-speaking peers in English (Hammer & Miccio, 2006; Hammer, Miccio, & Wagstaff, 2003; Páez et al., 2007; Tabors et al., 2003). Similarly, 78% of typically developing kindergarten-age ELs scored more than one standard deviation below monolingual norms on the Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001; Paradis, Schneider, & Sorensen Duncan, 2013), suggesting that difficulties in grammatical morphology are prevalent during the process of acquiring English as a second language (see Paradis, 2015 for an overview of typical and atypical development of English as a second language). At the same time, oral language proficiency is critical for reading comprehension (Geva, 2006), and oral language skills in the home language have been found to predict emergent literacy skills in English (Hammer, Lawrence, et al., 2007; Hammer, Rodriguez, et al., 2007; Rinaldi & Páez, 2008).

Language-specific characteristics, such as the transparency of phoneme-grapheme associations, account for variance in bilingual children's literacy skills. Spanish has relatively transparent orthography, with almost one-to-one mapping between graphemes and phonemes, whereas English has more opaque orthography. Thus, the opaque nature of English orthography

reduces the predictability of word pronunciations based on regularized phonological rules more than would be possible in a transparent language, such as Spanish (Devonshire, Morris & Fluck, 2013). For instance, by recognizing that the Spanish words *pato* (duck) and *gato* (cat) rhyme, a student can deduce that they are spelled similarly except for the initial letter. However, using the same assumption for the English words *share* and *there*, a spelling error is more likely to occur. Hence, it is likely that bilingual children draw upon their linguistic knowledge in Spanish to deepen their phonological awareness in English, which is corroborated by findings that suggest Spanish phonological and orthographic processing skills predict English reading skills, but not English spelling skills (Sun-Alperin & Wang, 2009). Studies that investigated the potential benefit of cross-linguistic transfer in the sematic domain found that vocabulary skills in the home language seem to facilitate the acquisition of cognates in English (Kohnert & Derr, 2004; Méndez-Pérez, Peña, & Bedore, 2010). A common denominator in this growing body of evidence is that the studies are either restricted in terms of sample size or they tend to be cross-sectional in their design, both of which temper their developmental implications.

For the development of literacy skills in bilingual children, oral language might have a unique status in that oral language proficiency might be an even stronger predictor of reading achievement than it is for monolingual children. Research has shown that the reasons for low reading performance in bilingual children do not lie in students' basic word-level skills (decoding, word recognition and spelling), but rather in students' text-level skills (August & Shanahan, 2006; Lesaux, Crosson, Kieffer, & Pierce, 2010). Those difficulties in reading comprehension were attributed to students' oral English proficiency. However, ELs' oral language skills are distributed across two languages, so that only focusing on oral English

proficiency does not take into account the dynamic nature of bilingual children´s language profiles.

**Instructional Programs for ELs**

Another dimension along which ELs educational experiences differ from those of their monolingual peers is the linguistic diversity of the instructional programs in which they participate, and the extent to which their primary language is or is not used instructionally. Children designated as ELs by their respective school districts are placed in one of several programs of instruction that vary with respect to the degree to which students' primary language is involved in instruction, which is an important contextual factor to consider in any study of oral language and literacy development in school-age children. Among the myriad of programs of instruction, some of the most common across the country are: Structured English immersion (SEI), transitional bilingual education (TBE), and dual language programs (Swanson, 2009). Each of these distinct programs has specific goals with regards to developing literacy skills only in English (SEI), using the primary language to bridge to English (TBE), or developing literacy in English and the student's first language (dual language; developmental bilingual education, and maintenance programs) (August & Shanahan, 2006).

Instructional contexts for language minority students vary across the developing world outside the US. In Canada, heritage language programs are prevalent for French speaking students, as well as dual language programs that focus on English and French for students from French and English speaking backgrounds. According to data from PISA 2015, most language minority students in EU countries receive little support for instruction in the primary language, although countries differ in their policies related to heritage language learning and instruction. For example, Sweden and Norway present exceptions to the general rule among EU school

systems in so far as they provide students the right to learn in their heritage language (Eurydice, 2004 as reported in Haahr, Nielsen, Hansen, & Jakobsen, 2005).

It is important to note that the purpose of the analyses reported here is not to demonstrate whether one specific program of instruction is better than another, but to characterize and describe the oral language and reading development of Spanish-speaking ELs in the US considering a range of factors (e.g., language of instruction) that may potentially contribute to that characterization. It is arguably preferable to employ longitudinal data when examining the development of educational outcomes in students whose mastery of the societal language is changing in ways and at a rate that may not be well-described by normative trajectories for the monolingual population (Mancilla-Martinez & Lesaux, 2011). Systematically identifying factors that contribute to variance in ELs performance across time in both the native language and English helps to simultaneously characterize both the rich heterogeneity of this population and their emerging development of both languages.

We systematically investigate this heterogeneity by (a) tracking bilingual children's oral language and reading skills across both languages from kindergarten to second grade using a comprehensive test battery, and (b) examining sources of variability in student test scores over time, including the language of instruction. We address various limitations of previous studies by investigating an exceptionally large longitudinal sample (over 4,100 participants) over 3 academic years, which improves generalizability of findings and captures a more accurate developmental picture of a broad range of language and literacy skills in English and Spanish. Assessing both written and spoken language comprehension skills acknowledges the strong relationship between these two domains, as early oral language skills are predictive of later

reading abilities in monolingual English-speaking and Spanish-English bilingual children (Catts, Fey, Tomblin, & Zhang, 2002; Hammer et al., 2014; Miller et al., 2006).

## Method

The present investigation utilized existing data from two longitudinal, large-scale parent projects: the *Oracy/Literacy Development of Spanish-Speaking Children* (Francis et al., 2005) (OLDS) project, and the *Success through Academic Interventions in Language and Literacy* (SAILL) project. Fifty-one schools and 12 districts were involved across both projects. OLDS involved cross-sectional and longitudinal studies of Spanish-speaking ELs in early elementary school (K through Grade 3) in large urban districts and in smaller districts in different regions of the US Only the longitudinal study is considered here. That study selected students in K in various bilingual programs and in structured English immersion programs and followed from the beginning of K through the end of Grade 2. The research was structured to examine factors that impacted developmental trajectories in literacy and oral language skills in English and Spanish and included an initial sample of 1,448 Spanish-speaking students, who were enrolled during kindergarten in California (Los Angeles and Orange County areas) and Texas (Austin, Houston, and Brownsville). The initial sample of EL students was expanded in each additional year of the project (Grades 1 and 2) to offset attrition. The final sample from the OLDS longitudinal project included a total of 2,109 EL students. Inclusionary criteria required that participants were categorized as EL students by district-based language proficiency testing. They attended a range of programs of instruction (structured English immersion (SEI); transitional bilingual education (TBE); dual language; Spanish maintenance bilingual program) and had never received special education services.

SAILL used a similar design to the OLDS longitudinal project and tracked Spanish-speaking ELs longitudinally from kindergarten to third grade in a district along the US-Mexico border to test the effectiveness of two research-enhanced models of instruction (SEI; TBE) on literacy and language development in English and Spanish. This longitudinal project initially included 1,202 ELs enrolled during K in Brownsville, Texas. The initial sample of students was expanded in subsequent years of the project (grades 1 and 2) to offset attrition. The final sample from the SAILL longitudinal project included 2,025 EL students; inclusionary criteria paralleled the OLDS longitudinal project. The use of parallel longitudinal designs and assessment protocols in both English and Spanish made it possible to combine the data from these projects to examine the factors affecting student performance that may influence the process of identification of students with disabilities. In the present study, we concentrate exclusively on the combined longitudinal samples from the SAILL and OLDS projects from the beginning of K to the end of Grade 2.

The Woodcock Language Proficiency Battery-Revised (WLPB-R; Woodcock, 1991; Woodcock & Muñoz-Sandoval, 1995) was administered to each student in the fall and spring semesters of each academic year in Spanish and English, involving 6 waves of data collection between the fall of K and the spring of grade 2, with administration of the WLPB-R toward the beginning of the fall semesters, and late spring. Participants were first administered the WLPB-R in their strongest language as determined by the examiners through informal interaction with each child preceding testing; they were administered the WLPB-R in their other language approximately one week later by trained examiners who were proficient speakers of both languages. Every child with parental consent in every classroom who met the inclusionary criteria was assessed throughout the duration of each project. Correspondingly, children were

added to the sample in Grade 1 and Grade 2 at the start of the year to augment the sample due to year-over-year attrition. Although measures were generally collected in the fall and spring of each year, there were some exceptions: we did not measure reading comprehension in K in either project, and SAILL did not measure phonological awareness in Grade 2. Thus, the number of waves of participation varied across children due to when they were recruited, attrition, and missed assessments, and the number of observations per measure varied depending on these same factors as well as on the aforementioned differences in the planned assessment schedules for some measures. We used a systematic merging procedure in order to construct one final longitudinal dataset that included data from all ELs who completed assessments as part of the two longitudinal projects. Each student's data are included for every valid assessment completed by the student. It was not necessary that the student have completed all assessments in a given language, all assessments in both languages, nor assessments at all time points for their data to be included in the analyses.

The current study's final dataset included a total of 15,407 observations from 4,134 ELs assessed on subtests of the WLPB-R measuring the Broad Reading Cluster and Oral Language Cluster in English and Spanish in the fall and spring semesters from K through grade 2. Table 1 summarizes the overall sample size across the two projects and at each time point, participant gender, and mean performances in the fall and spring on the WLPB-R outcome measures (described below) in each language for each project. There were slightly more observations for boys (51%; $n$ = 7,886) than girls (49%; $n$ = 7,521) in the final dataset.

**Outcome Measures**

The present study examined sources of variability in the oral language and reading skills of Spanish-speaking ELs throughout the early elementary school grades. We focus on the age

standardized scores at each grade and wave, as these are fundamental to the identification of disabilities using several approaches. Other papers in the volume may use the W score in their analyses. The W score is appropriate for modeling change over time and is constructed to allow both the mean and variance to change over time with respect to development whereas the age standard score is constructed to have a stable mean and variance at each standardization point in the norming population. Thus, changes in the mean and variance of performance within the EL student population reflects performance gains or losses and changes in score dispersion relative to norms. These differences can be informative for understanding the effects of instruction and development within this population of students, who are often economically disadvantaged (McFarland et al., 2017; OECD 2006).

The Broad Reading Cluster of the WLPB-R in Spanish and English is a broad measure of reading achievement. This cluster is a combination of measures of reading identification, comprehension, and vocabulary skills. This cluster consists of two subtests: Letter-Word Identification, and Passage Comprehension. The Letter-Word Identification subtest measures the child's ability to identify letter forms or words they have not seen before. The first five items measure the child's ability to match a pictographic representation of a word with an actual picture of an object. The remaining items measure the child's ability to first identify letters, which are presented in large type, and then to pronounce the presented word correctly. The items in this test become more difficult as the child is presented words that are used less frequently in the written language (English or Spanish). The Passage Comprehension subtest measures the child's comprehension and vocabulary skills. The first four items in this test are presented in a multiple-choice format; the child points to the picture represented by a phrase. The remaining

items measure the child's skill in reading a short passage and identifying a missing key word; the child is required to state the word that is appropriate in the context of the passage.

The Oral Language Cluster of the WLPB-R in Spanish and English is a measure of linguistic competency, semantic expression, expressive vocabulary, and listening and verbal comprehension. This cluster is comprised of five subtests: Memory for Sentences, Picture Vocabulary, Oral Vocabulary, Listening Comprehension, and Verbal Analogies. The Memory for Sentences subtest measures the child's ability to repeat, verbatim, phrases and sentences of increasing length and syntactic and semantic complexity that have been presented auditorily. The Picture Vocabulary subtest measures the child's expressive vocabulary. Students are asked to name pictured objects that reflect single word meanings. The items are presented in order of increasing difficulty. The Oral Vocabulary subtest measures knowledge of synonyms and antonyms. The Listening Comprehension subtest measures the child's ability to comprehend a passage by supplying a single word missing at the end in an oral cloze procedure. The test is based on a number of semantic operations. It starts with simple verbal analogies and associations and progresses to a higher level of comprehension (i.e., the ability to discern implications). Finally, the Verbal Analogies subtest measures the child's ability to comprehend and verbally complete a logical word relationship. The relationship between the words starts from very simple and becomes increasingly complex.

Subtests of the WLPB-R are broadly recognized for their excellent psychometric properties, and are widely used in educational research, assessment of language proficiency and educational achievement, and educational diagnostic practice. Reliability and validity of the battery for use with Spanish-speaking children is well documented in the technical manuals, and based on data from the present samples. Internal consistency measures of reliability for the

present study were uniformly above .85 for all subtests in each language; intra-battery correlations among subtests for the reading and oral language subtests were uniformly strong. Specifically, for students instructed in English, reading subtest correlations ranged from .70 to .83 in English and from .66 to .82 in Spanish and oral language subtests correlations ranged from .48 to .69 in English and from .59 to .87 in Spanish. For students instructed in Spanish, reading subtest correlations ranged from .69 to .79 in English and from .72 to .89 in Spanish and oral language subtest correlations ranged from .55 to .75 in English and from .46 to .69 in Spanish. On balance, these correlations show evidence of both reliability and validity in Spanish-speaking ELs in both Spanish and English instruction.

**Analytic Approach**

We present two sets of multivariate multi-level models estimated using restricted maximum likelihood in PROC MIXED in SAS 9.4. We fit five models that differ in their complexity with respect to the variance components and mean structure. The models increase in complexity from Model 1 to Model 4. Model 5 is intermediate in complexity between Models 3 and 4. Models 1 and 2 estimated separate means for English and Spanish outcomes, but differ in their variance components. Model 1 estimated the total variance in outcomes, whereas Model 2 decomposes this total variance into variance at the school-, teacher-, and student-levels. Model 3 introduces a more complex mean structure by allowing the means in each language to differ between students in English and Spanish instruction. Thus, some of the variance in Models 1 and 2 is due to their failure to allow for mean differences between English and Spanish instruction. Models 4 and 5 retain the more complex mean structure of Model 3 and allow the variances in outcomes and their covariance at both the student and teacher levels to differ between English and Spanish instruction. Model 4 allowed for variance heterogeneity at the

classroom- and student-levels, and Model 5 allow for this heterogeneity only at the classroom-level.

In summary, the model is univariate with respect to time and outcome (reading or language), but multivariate with respect to the outcome language, such that at all time points we model one outcome in two languages simultaneously. In K, we model oral language, and in grades 1 and 2 we model reading and oral language outcomes; we model the two outcomes separately, but in both languages simultaneously. This approach was deemed necessary due to difficulties in estimating models in grades 1 and 2 that used all four outcomes simultaneously. Although all four outcomes could be analysed simultaneously using PROC HPMIXED, variance component estimates differed substantially between PROC MIXED and PROC HPMIXED for the same outcomes. Models with all four outcomes proved impractical to estimate using PROC MIXED; individual models required more than 48 hours to complete a single likelihood iteration, whereas models with two outcomes varied in time to estimate from a few minutes to approximately two hours on a Data Science Workstation with 96 GB of RAM and Intel i9-7900X – Skylake 10-Core hyperthreaded CPU @3.30GHz.

## Results

### Descriptive Analyses

Table 1 provides descriptive statistics and sample sizes for each of the subtests of interest in both English and Spanish across the six different time points. Table 1 highlights that children are performing substantially below normative expectations, especially in oral language in both English and in Spanish, and for the most part are performing below expectations in reading in both languages as well, albeit to a lesser extent. At the same time, we can see that students are gaining substantially against national norms over time in both projects. Specifically, in the

OLDS project, English oral language improves from a mean of 55.0 in the fall of K to a mean of 75.6 by the spring of grade 2, whereas Spanish oral language outcomes improved from a mean of 75.1 to 85.1. An improvement of 12.1 standard score points was observed in English oral language in Project SAILL, whereas Spanish oral language improved slightly from a mean of 68.2 to 72.9. Gains in reading were more dramatic in both projects and languages, although in this case gains are measured from the fall of grade 1 to the spring of grade 2, because reading was not measured in K. In English, reading scores improved from a mean of 90.9 to a mean of 103.2 in the OLDS project and from 85.8 to 99.2 in SAILL. In Spanish, reading improved to 110.6 from a mean of 100.2 in OLDS and to 101.7 from 93.1 in SAILL.

These improvements over time do not take into consideration any differences that might relate to the language of instruction, and do not give a sense of the substantial spread in performance across students. Figure 1 provides side-by-side box plots of performance in English and Spanish for oral language. The figure is organized into four panels in a 2 X 2 grid. The left-hand column presents the distributions for English (top) and Spanish (bottom) ignoring the language of instruction, whereas the right-hand column presents the distributions taking into account the language of instruction. The left hand panels contain six side-by-side box plots representing the six waves of data from left to right, beginning with the fall of K and ending with the spring of grade 2. The right hand panels show twelve side-by-side box plots showing the same six waves of data, but broken out for students in Spanish instruction (first six boxes) and students in English instruction (second six boxes).

These side-by-side box plots highlight the substantial variation associated with the language of instruction, and how these effects differ over time across the language of instruction and the language of the outcome. Children in English instruction begin K with performance in

English oral language that is quite similar to the English oral language performance of children in Spanish instruction at the end of grade 2. That is, we find the distribution of English proficiency in K for children instructed in English to be quite comparable to the distribution of English proficiency at the end of grade 2 for children instructed in Spanish. This comparability applies to both the mean performance as well as to the general spread in performance. The critical point is that the English language proficiency skills of children instructed in English and children instructed in Spanish are quite different ***at the start of K***. This difference cannot be the result of instruction or the language of instruction. Rather, it reflects a difference in English proficiency between the instructional groups that predates the onset of instruction. In contrast, children in Spanish instruction begin K with substantially higher Spanish oral language proficiency than children in English instruction. Moreover, even by the end of grade 2, the Spanish language skills of children in English instruction lag behind that of their Spanish-instructed peers at the start of K. Spanish language differences between the two instructional groups are large. In fact, the boxes that represent the middle 50% of the distributions are non-overlapping, indicating that 75% of the distribution of Spanish performance for students in Spanish instruction lies above 75% of the distribution of Spanish performance for student in English instruction. Like English, these differences in Spanish proficiency predate the onset of instruction in K, but in this case the difference favours students in Spanish instruction.

Figure 1 makes clear that substantial variation in performance in both English and Spanish is associated with the organization of children into Spanish and English instruction, but also that the magnitude of these organizational effects is not the same across English and Spanish outcomes. The existence of these differences at the start of each school year signals that they are

likely to be organizational effects related to the sorting of children into instructional groups, as well as instructional effects that unfold over time.

It is readily apparent from Figure 1 that the standard score distributions are not stable over time in the same way that one might expect for monolingual students. These differences are not the result of cohort effects or attrition. Importantly, it is clear from comparing the right- and left- hand panels that any low achievement criterion in English that ignores language of instruction will over-identify students in Spanish instruction. Moreover, if a criterion is established in English based only on the performance of EL students, that criterion will tend to under-identify students in English instruction. Based on national norms for monolingual students, over-identification is likely regardless of the language of instruction, and classifications are likely to be wildly unstable over time as children's performance improves against national norms. (See Santi et al., this volume for an empirical example).

**Analysis of Multi-level Models**

Also apparent from Figure 1 is the substantial variability in performance in both languages, in both instructional groups, and at each measurement occasion. Some of this variability exists at the student-level, the classroom-level, and the school-level. To better capture these sources of variability, we fit a series of multi-level models to the multivariate data at each wave for reading and oral language outcomes as described above.

Table 2 presents variance component estimates for Model 4 for each of the six occasions of measurement for both sets of outcomes. We present the variance components estimates for Model 4 because for both sets of outcomes and for all occasions of measurement, Model 4 provided the best fit to the data based on the likelihood ratio, the AIC, BIC, and AICC. Thus, Model 4 was clearly superior to the alternative models for all six occasions of measurement and

both sets of outcomes. Model 4 allowed for variability at the student, teacher, and school levels, with common variance components at the school level, but separate variance components at the teacher and student levels for English and Spanish instruction, and separate mean vectors for the two kinds of instruction. For all occasions of measurement, Model 3 provides a substantially better fit than Model 2 and Model 1, which implies that the mean structure must take into account the language of instruction and the variance structure must distinguish between students, teachers, and schools. Model 5 constrained the student-level residual variance to be the same for English and Spanish instruction, but this constraint introduced a lack of fit relative to Model 4. Thus, on the basis of the model fit statistics, we conclude that at each occasion of measurement there is statistically significant variance at the school level, and substantial variation at the teacher- and student-levels that differs across English and Spanish instructional groups. At the same time, the mean vectors differ across English and Spanish instructional groups. (Fit statistics for all 50 models – five models over six occasions for oral language and 5 models over four occasions for reading – are available upon request from the first author).

Examination of Table 2 shows that substantial variation exists at all three levels of the design, with most of the variability at the student level. Perhaps most interesting in Table 2 are the magnitude and direction of the correlation between English and Spanish outcomes at the school-, teacher-, and student-levels and how these vary across occasions of measurement. At the student level, correlations are uniformly positive, but vary in magnitude depending on the outcome (oral language or reading), the time point, and the type of instruction. Specifically, student-level correlations are larger for reading than for oral language, suggesting that reading skills are more highly related across languages at the student level in these early grades than oral language skills. Moreover, the reading correlation is more consistent across language of

instruction at the student level than is the correlation between English and Spanish oral language skills. Moreover, in the latter grades, language skills seem to become less related for students in English instruction compared to students in Spanish instruction, although the correlation is not strong in either type of instruction.

In contrast to the student-level correlations, the correlations at the classroom- and school-levels are not uniformly positive. Classroom correlations in oral language are generally negative, and occasionally quite strongly negative indicating that classrooms with stronger English performance tend to have lower Spanish performance and vice versa. Only in the spring of K in Spanish-instructed classrooms is the oral language correlation positive ($r$ = .14). In all other instances the correlation is negative at the classroom level ranging from -.24 to -.97. In general, the correlation is more negative at each grade and time point for English instruction than for Spanish instruction, indicating that oral language performance is less related across languages for bilingual classrooms and more strongly negatively related for English instruction classrooms. In reading, classroom correlations are positive in the fall of grade 1, as well as in the spring of grade 1 for Spanish instruction classrooms, but negative for English instruction classrooms. In grade 2, the correlation is negative for Spanish instruction in both the fall and spring, but positive for English instruction classrooms in the fall and negative in the spring. We view the fall correlations as reflecting organizational effects (i.e., how students are sorted and assigned to classrooms for instructional purposes), whereas the spring correlations reflect the effects of organization, and effects of instruction and how instruction shapes performance across the two languages. At the school-level, oral language correlations were negative in K, but positive in grades 1 and 2, whereas correlations between English and Spanish reading were positive, but decreased from the spring of grade one to the fall of grade 2.

To facilitate comparison across time points and levels in terms of the variance that resides at the student-, teacher-, and school-levels, we present the variance components of Table 2 as intra-class correlations in Table 3. To compute school-level intra-class correlations, the total variance was computed by pooling the Total under English and Spanish instruction. The information in Table 3 reveals that most of the variability in oral language and reading is at the student level, namely from 67% to 90% in oral language and from 67% to 94% in reading. Variance at the teacher level ranged from 4% to 27% in oral language and from 3% to 32% in reading, whereas variance at the school level ranged from 2% to 11% in oral language and from 6% to 11% in reading. In oral language, the variance at the teacher level appears to be greater in English as compared to Spanish instruction in K and grade 1 in both English and in Spanish outcomes, but appears comparable across types of instruction in grade 2 in both languages. Variance in reading outcomes at the classroom level is relatively comparable between English and Spanish instruction for English outcomes, but is substantially greater in Spanish as compared to English instruction for Spanish outcomes.

Figure 2 provides a plot of the means for oral language and reading outcomes in English and in Spanish for children in English and Spanish instruction. These figures showcase the interactive roles that language of outcome and language of instruction play in the evolution of performance in oral language and reading outcomes from the beginning of K to the end of second grade. Specifically, we can readily see that oral language performance falls below that of reading performance. Moreover, we see that oral language skill in Spanish is best for students in Spanish instruction, whereas oral language skill in English is best for students in English instruction, although English oral language performance is lower for students in English instruction than is Spanish oral language for students in Spanish instruction. We also see that

English oral language performance for students in Spanish instruction is below the Spanish oral language performance for students in English instruction, but this difference declines over time, such that students in Spanish instruction have begun to close the gap by the end of grade 1 and their English proficiency exceeds the Spanish proficiency of English instructed students by the spring of grade 2. This reversal stems from the fact that Spanish proficiency as measured by the standard scores changes little for students in English instruction from the fall of K through the spring of grade 2, whereas English proficiency begins to improve for children in Spanish instruction beginning in grade 1. The pattern of performance in reading is different in so far as English reading of Spanish instructed students exceeds the Spanish reading of English instructed students on average at the start of grade 1. Finally, the two panels of Figure 2 make clear that by the spring of grade 2 the greatest differences between instructional groups are in Spanish outcomes, both oral language and reading as compared to English outcomes.

## Discussion

The identification of disabilities among language minority students presents unique challenges for schools. On the one hand, societies acknowledge the importance of identifying and treating learning disabilities, and in the case of the US, federal law specifies that procedures used for identification among majority students can be used for identification of disabilities in language minority students, as well. However, guidance on how best to implement such procedures with language minority students is lacking, as is empirical evidence on the performance of any such methods. The present study utilized multivariate multilevel models to examine variability in oral language and reading outcomes for Spanish-speaking ELs in the US. Through these analyses the study highlights the extensive variability in student oral language and reading performance, that this variability exists in both the societal language and the students'

home language, and that language of instruction is a non-ignorable factor when considering expectations with regards to performance. Moreover, the findings highlighted that performance expectations vary with respect to outcome (reading or oral language), language (English or Spanish), occasion of measurement, and the language of instruction.

This variation in performance expectations has profound implications for identifying reading and language disabilities among language minority students. In the face of such variability, one cannot apply simple cut-offs based on societal normative performance, or even norms based on the target population that do not also take into account factors such as schooling context, especially the language of instruction. The study showed that substantial variability exists at the school- and classroom-level, even after the language of instruction is taken into consideration. In some cases, the variability at the classroom level exceeded 20% and even 30%. Moreover, variability at both the student- and classroom- levels was found to differ depending on the language of instruction.

Taken together these findings have implications for the identification of students with disabilities, but the exact implications are less than clear. Even something as apparently straight forward as controlling for the language of instruction is not clear-cut. In our study, we created two categories of instruction, English and Spanish, with the latter category including any approach that made use of students' home language. Thus, the "Spanish" category included transitional bilingual education, dual language instruction, and maintenance/developmental bilingual education programs. We did not examine the possibility that these programs might necessitate different expectations for students' English and Spanish oral language and literacy development. Moreover, students often spend time in different programs, either because they move from one school to another school, or because schools make decisions to change program

models collectively or for individual students. The implications that such programming adjustments have for student performance expectations were not considered here, but it seems likely that a child's full educational history with respect to language of instruction could impact expectations for oral language and literacy development. In so far as 65% to 85% of the variability in performance resided at the student level, after controlling for the language of instruction as defined by our program models, other potential child-level covariates could be incorporated into the models to improve the development of performance targets against which to gauge student's observed performance in order to determine whether a given student's development is on track, or not.

The models that we have put forward are not explanatory, but merely descriptive of the levels at which scores vary and by how much. This description points to the types of covariates that could be taken into account in setting expectations for student performance. The use of covariates in setting developmental expectations for the purposes of disability identification is akin to the use of local norms introduced in the literature on response to intervention **(**Shin, 1988). However, in that literature the focus is largely on the identification of students for targeted intervention, or more intensive instruction, not on the use of local norms for disability identification per se. Explicit use of covariates in disability identification carries substantial risk for limiting identification in school settings characterized by low performance, whereas the failure to consider contextual influences risks over identification of students in those settings and increases the demand for limited Special Education resources. Better longitudinal studies with more protracted follow-up could mitigate this dilemma by allowing the development of long term projections based on student characteristics and instructional experiences. Such extensive longitudinal data systems would allow the identification of students based on the student's

individual risk for long-term negative educational outcomes and then targeting specific educational resources to those children based on evidence of risk reduction associated with specific interventions/instructional experiences. The viability of such an approach may require that resources be allocated based on the risk profiles of students within schools rather than on explicit identification of students as disabled/not disabled.

**Limitations**

This study was made possible by the joining of multiple datasets collected for different purposes. Although the dataset is extensive for studies of this sort, it is also limited to Spanish-speaking ELs in a limited set of educational contexts in the US. Although Spanish-speaking ELs comprise the largest language group in the US, there are substantial numbers of ELs in the US context who speak other languages and who are receiving their education under very different school contexts. Indeed, the most common approach to instruction for ELs in the US is structured English immersion, but native language instruction was most prevalent in our studies. As discussed in the introduction, language minority students are not unique to the US school context and in fact represent substantial portions of the student body throughout the developed world (OECD, 2106). Extension of this study to other US contexts and to non-US contexts is important and pressing, but possibly demands a scale of investment that cannot be achieved through investigator-initiated research. Of course, most important are the development of effective interventions and the allocation of requisite resources to deliver those interventions to those students in need. At the same time, broader study of the identification of disabilities among language minority students would serve to highlight the scope of the problem and its associated challenges for researchers and practitioners.

The focus on a single language group is a major limitation of the current study. Our study confounds language minority status with the specific language spoken by our students, as well as other background characteristics that differentiate our students from the societal mainstream. In the case of Spanish-speaking ELs in the US, language group is confounded with socioeconomic status as well as the educational level of the parents (McFarland, et al., 2017). This confounding makes it impossible to disentangle the potentially unique effects of students' primary language, parental education, and socioeconomic status on student performance. Broadening the sample to include speakers of other alphabetic languages as well as non-alphabetic languages, such as Mandarin and Cantonese, would strengthen this research and improve our ability to disentangle these factors and determine their implications for identification of disabilities in language minority students. At the same time, broadening the language representation in the study creates additional complications for obtaining commensurate measures in students' L1's. Nevertheless, these challenges need to be addressed and the research extended in order to improve identification and treatment of disabilities in language minority children.

## References

Anthony, J. L., Solari, E. J., Williams, J. M., Schoger, K. D., Zhang, Z., Branum-Martin, L., et al. (2009). Development of bilingual phonological awareness in Spanish- speaking English language learners: The roles of vocabulary, letter knowledge, and prior phonological awareness. *Scientific Studies of Reading, 13,* 535–564.

August, D., & Shanahan, T. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on language-minority children and youth*. Mahwah, NJ: Erlbaum.

Basterra, M. R., Trumbull, E., & Solano-Flores, G. (2010). *Cultural validity in assessment: Address linguistic and cultural diversity.* New York, NY: Routledge.

Branum-Martin, L. Tao, S., Garnaat, S., Bunta, F., & Francis, D.J. (2012). Meta-Analysis of Bilingual Phonological Awareness: Language, Age, and Psycholinguistic Grain Size. *Journal of Educational Psychology, 104(4),* 932-944.

Catts, H., Fey, M., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142– 1157.

Devonshire, V., Morris, P., & Fluck, M. (2013). Spelling and reading development: The effect of teaching children multiple levels of representation in their orthography. *Learning and Instruction, 25*, 85–94.

Francis, D., Carlson, C., Fletcher, J., Foorman, B., Goldenberg, C., Vaughn, S.,…Papanicolaou, A. (2005). Oracy/Literacy Development of Spanish-speaking Children: A multi-level program of research on language minority children and the instruction, school and

community contexts, and interventions that influence their academic outcomes. *Perspectives: The International Dyslexia Association, 31,* 8-12.

Geva, E. (2006). Second language oral proficiency and second language literacy. In D. August, & T. Shanahan (Eds.), *Developing literacy in second-language learners: A report of the national literacy panel on language minority children and youth* (pp. 123–139). Mahwah, NJ: Erlbaum.

Geva, E., & Wang, M. (2001). The development of basic reading skills in children: a cross-language perspective. *Annual Review of Applied Linguistics, 81,* 182-204.

Haahr, J., Nielsen, T.K., Hansen, M.E., & Jakobsen, S.T. (2005). Explaining student performance: Evidence from the international PISA, TIMSS, and PIRLS surveys. www.danishtechnology.dk.

Halle, T., Hair, E., Wandener, L., McNamara, M., & Chien, N. (2012). Predictors and outcomes of early versus later English language proficiency. *Early Childhood Quarterly, 27,* 1–20.

Hammer, C. S., Lawrence, F. R., & Miccio, A. W. (2007). Bilingual children's language abilities and reading outcomes in Head Start and kindergarten. *Language, Speech and Hearing Services in Schools, 38,* 237–248.

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D. C., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*(4), 715–733.

Hammer, C. S., & Miccio, A. W. (2006). Early language and reading development of bilingual preschoolers from low-income families. *Topics in Language Disorders, 26,* 302–317.

Hammer, C. S., Miccio, A. W., & Wagstaff, D. (2003). Home literacy experiences and their relationship to bilingual preschoolers' developing English literacy abilities. *Language Speech and Hearing Services in Schools, 34,* 20–30.

Hammer, C. S., Rodriguez, B. L., Lawrence, F. R., & Miccio, A. W. (2007). Puerto Rican mothers' beliefs and home literacy practices. *Language, Speech, and Hearing Services in Schools, 38,* 216–224

Han, W. (2012). Bilingualism and academic achievement. *Child Development, 83,* 300–321

Hoff, E. (2013). Interpreting early language trajectories of children from low-SES and language minority homes: Implications for closing the achievement gap. *Developmental Psychology, 49*(1), 4-14.

Kohnert, K., Derr, A. (2004). *Language intervention with bilingual children*. In: Goldstein, BA., (ed.) *Bilingual language development and disorders in Spanish-English speakers.* Baltimore, MD: Brookes, 311-338.

Lesaux, N., Crosson, A., Kieffer, M.J., & Pierce, M. (2010). Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. *Journal of Applied Developmental Psychology, 31(6),* 475-483.

Mancilla-Martinez, J., & Lesaux, N. (2011). The gap between Spanish-Speakers' word reading and word knowledge: A longitudinal study. *Child Development, 82*, 1544-1560.

McFarland, J., Hussar, B., de Brey, C., Snyder, T., Wang, X., Wilkinson-Flicker, S., . . . Hinz, S. (2017). *The condition of education 2017 (NCES 2017-144).* Washington, DC: US Department of Education, National Center for Education Statistics.

Méndez-Pérez, A., Peña, E., Bedore, L. (2010). Cognates Facilitate Word Recognition in Young Spanish-English Bilinguals´ Test Performance. *Early Child Services, 4 (1),* 55-67.

Miller, J., Heilmann, J., Nockerts, A., Iglesias, A, Fabiano, L., & Francis, D. (2006). Oral language and reading in bilingual children. Learning Disabilities Research and Practice, *21*, 30-43.

National Center for Education Statistics. (2015). *Digest of education statistics.* Washington, DC: US Department of Education.

OECD (2016), PISA 2015 Results (Volume I): Excellence and Equity in Education, PISA, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264266490-en

Oller, D.K.& Eilers, R.E. (eds) (2002). *Language and Literacy in Bilingual Children.* Clevedon, UK: Multilingual Matters.

Páez, M. M., Tabors, P. O., & López, L. M. (2007). Dual language and literacy development of Spanish-speaking preschool children. Journal of Applied Developmental Psychology, 28,85–102.

Paradis, J. (2015). The development of English as a second language with and without specific language impairment: Clinical implications. *Journal of Speech, Language and Hearing Research, 59,* 171-182.

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language and Hearing Research, 56,* 971–981.

Rice, M., & Wexler, K. (2001). *Test of Early Grammatical Impairment*. New York: The Psychological Corporation.

Rinaldi, C. & Páez, M. (2008). Preschool matters: Predicting reading difficulties for Spanish-speaking students in first grade. *Learning Disabilities: A Contemporary Journal, 6(1),* 71-84.

Saunders, W.M. & Marcelletti, D.J. (2013). The gap that can't go away: The catch-22 of reclassification in monitoring the progress of English Learners. *Educational Evaluation and Policy Analysis, 35(2),* 139-156.

Shin, M. (1988). Development of curriculum-based local norms for use in Special Education. *School Psychology Review, 17(1),* 61-80.

Sun-Alperin, M., & Wang, M. (2011). Cross-language transfer of phonological and orthographic processing skills from Spanish L1 to English L2. *Reading and Writing, 24(5),* 591-614.

Swanson, C. B. (2009). *Perspectives on a population: English-language learners in American schools*. Bethesda, MD: Editorial Projects in Education.

Tabors, P. O., Páez, M. M., & López, L. M. (2003). Dual language abilities of Spanish-English bilingual four-year-olds: Initial findings from the early childhood study of language and literacy development of Spanish-speaking children. *NABE Journal of Research and Practice, 1,* 70–91.

Umansky, I.M., Thompson, K.D., & Diaz, G. (2017). Using an ever–English Learner framework to examine disproportionality in Special Education. *Exceptional Children, 84(1),* 76-96.

Woodcock, R. W. (1991). *Woodcock Language Proficiency Battery-Revised*. Itasca, IL: Riverside Publishing.

Woodcock, R. W., & Muñoz-Sandoval, A. F. (1995). *Woodcock Language Proficiency Battery-Revised: Spanish Form*. Itasca, IL: Riverside Publishing.

Table 1. Descriptive Statistics broken down by Project

| | K-Fall | K-Spring | 1st-Fall | 1st-Spring | 2nd-Fall | 2nd-Spring |
|---|---|---|---|---|---|---|
| OLDS Project: Participants | 1,464 | 1,348 | 1,628 | 1,507 | 1,380 | 1,258 |
| Classrooms (S;E) | 65;35 | 67;36 | 113;76 | 112;73 | 104;67 | 105;66 |
| Boys; Girls | 736; 728 | 680; 668 | 851; 777 | 776; 731 | 695; 685 | 636; 622 |
| WLB: BRE | NC | NC | 90.9 (19.2) | 98.5 (19.4) | 102.8 (16.8) | 103.2 (16.8) |
| WLB: OLE | 55.0 (19.8) | 59.5 (20.7) | 59.3 (21.9) | 64.6 (20.7) | 70.4 (18.6) | 75.6 (17.9) |
| WLB: BRS | NC | NC | 100.2 (28.1) | 110.6 (28.1) | 113.3 (25.4) | 110.6 (27.1) |
| WLB: OLS | 75.1 (22.1) | 80.3 (23.2) | 80.9 (23.4) | 83.0 (24.6) | 84.5 (23.7) | 85.1 (24.4) |
| SAILL Project: Participants | 1,202 | 1,155 | 1,185 | 1,059 | 1,156 | 1,065 |
| Classrooms (S;E) | 36;19 | 36;19 | 44;21 | 44;21 | 39;23 | 39;23 |
| Boys; Girls | 625; 577 | 606; 549 | 611; 574 | 545; 514 | 584; 572 | 541; 524 |
| WLB: BRE | NC | NC | 85.8 (20.7) | 96.0 (21.3) | 95.9 (20.4) | 99.2 (19.7) |
| WLB: OLE | 57.6 (21.2) | 61.5 (22.9) | 57.4 (23.5) | 61.5 (23.1) | 64.8 (20.8) | 69.7 (19.8) |
| WLB: BRS | NC | NC | 93.1 (26.4) | 102.9 (28.4) | 102.5 (28.5) | 101.7 (27.9) |
| WLB: OLS | 68.2 (25.8) | 73.7 (26.5) | 70.0 (26.6) | 71.3 (27.1) | 71.3 (27.6) | 72.9 (28.1) |

*Note*. K = kindergarten; 1st = first grade; 2nd = second grade; OLDS = *Oracy/Literacy Development of Spanish-Speaking Children*; WLB = Woodcock Language Proficiency Battery-Revised, Age Standard Score (Woodcock, 1991; Woodcock & Muñoz-Sandoval, 1995); BRE = Broad Reading Cluster in English; OLE = Oral Language Cluster in English; BRS = Broad Reading Cluster in Spanish; OLS = Oral Language Cluster in Spanish; SAILL = *Success through Academic Interventions in Language and Literacy*; NC = Not collected; Standard deviations appear in parentheses.

Table 2. Variance components estimates from Model 4 for each of six occasions of measurement.

| | | Variance Estimates | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Kindergarten Fall Estimate | | Kindergarten Spring Estimate | | Grade 1 Fall Estimate | | Grade 1 Spring Estimate | | Grade 2 Fall Estimate | | Grade 2 Spring Estimate | |
| Level | Measure | | | | | | | | | | | | |
| School | OLETC | 32.21 | | 38.90 | | 27.07 | | 26.15 | | 16.38 | | 13.63 | |
| | OLSTC | 32.93 | | 47.89 | | 10.19 | | 15.67 | | 33.05 | | 32.28 | |
| | Corr. | -0.43 | | -0.74 | | 0.47 | | 0.60 | | 0.36 | | 0.44 | |
| | BRETC | NA | | | | 21.41 | | 35.63 | | 28.40 | | 18.26 | |
| | BRSTC | | | | | 54.67 | | 34.89 | | 58.42 | | 58.69 | |
| | Corr. | | | | | 0.55 | | 0.22 | | 0.21 | | 0.05 | |
| | | Instruction | | Instruction | | Instruction | | Instruction | | Instruction | | Instruction | |
| | | Spanish | English | Spanish | English | Spanish | English | Spanish | English | Spanish | English | Spanish | English |
| Teacher | OLETC | 22.45 | 39.20 | 27.40 | 56.53 | 35.65 | 104.76 | 39.92 | 79.83 | 88.13 | 45.80 | 86.75 | 61.63 |
| | OLSTC | 17.05 | 128.32 | 28.70 | 99.17 | 29.69 | 124.23 | 50.05 | 139.19 | 98.95 | 135.12 | 101.20 | 125.61 |
| | Corr. | -0.29 | -0.92 | 0.14 | -0.90 | -0.43 | -0.88 | -0.24 | -0.97 | -0.51 | -0.62 | -0.53 | -0.73 |
| | BRETC | NA | | | | 23.30 | 29.45 | 22.60 | 43.35 | 40.59 | 37.70 | 56.42 | 45.25 |
| | BRSTC | | | | | 61.18 | 20.33 | 100.30 | 59.96 | 105.62 | 44.29 | 141.78 | 68.25 |
| | Corr. | | | | | 0.36 | 0.30 | 0.28 | -0.23 | -0.10 | 0.26 | -0.18 | -0.12 |
| Student | OLETC | 299.11 | 275.18 | 301.47 | 252.31 | 352.65 | 263.00 | 313.73 | 231.44 | 255.91 | 205.57 | 236.60 | 176.42 |
| | OLSTC | 349.69 | 456.04 | 342.56 | 470.49 | 360.91 | 518.30 | 330.35 | 521.14 | 324.68 | 511.27 | 310.93 | 551.06 |
| | Corr. | 0.08 | 0.12 | 0.10 | 0.16 | 0.13 | 0.13 | 0.19 | 0.12 | 0.12 | 0.02 | 0.13 | 0.04 |
| | BRETC | 327.37 | 210.43 | 349.15 | 210.15 | 285.92 | 221.97 | 272.07 | 212.08 | 327.37 | 210.43 | 349.15 | 210.15 |
| | BRSTC | 537.96 | 382.68 | 369.10 | 427.81 | 359.41 | 433.38 | 358.19 | 474.70 | 537.96 | 382.68 | 369.10 | 427.81 |
| | Corr. | 0.72 | 0.57 | 0.64 | 0.58 | 0.61 | 0.61 | 0.61 | 0.56 | 0.72 | 0.57 | 0.64 | 0.58 |
| Total | OLETC | 353.77 | 346.60 | 367.77 | 347.74 | 415.36 | 394.83 | 379.81 | 337.43 | 360.42 | 267.75 | 336.98 | 251.68 |
| | OLSTC | 399.67 | 617.29 | 419.15 | 617.56 | 400.79 | 652.72 | 396.07 | 676.00 | 456.68 | 679.44 | 444.41 | 708.95 |
| | BRETC | NA | | | | 397.36 | 313.86 | 371.96 | 310.42 | 324.89 | 271.66 | 311.29 | 239.94 |
| | BRSTC | | | | | 653.82 | 457.68 | 504.29 | 522.67 | 523.45 | 536.09 | 558.66 | 601.63 |

NOTE: OLETC – Woodcock Oral Language composite in English; OLSTC – Woodcock Oral Language Composite in Spanish;

BRETC – Woodcock Broad Reading cluster in English; BRSTC – Woodcock Broad Reading Cluster in Spanish; NA - Reading was

not assessed in Kindergarten. Variance at the student and teacher levels is estimated separately for Spanish and English instruction in Model 4, whereas school level variance is not estimated separately for Spanish and English instruction.

Table 3. Intra-class Correlations for Model 3 for Reading and Language outcomes in English and Spanish from Kindergarten through Grade 2.

| | | Variance Estimates | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Level | Measure | Kindergarten Fall Estimate | | Kindergarten Spring Estimate | | Grade 1 Fall Estimate | | Grade 1 Spring Estimate | | Grade 2 Fall Estimate | | Grade 2 Spring Estimate | |
| School | OLETC | 0.09 | | 0.11 | | 0.07 | | 0.07 | | 0.05 | | 0.05 | |
| | OLSTC | 0.06 | | 0.09 | | 0.02 | | 0.03 | | 0.06 | | 0.06 | |
| | BRETC | NA | | | | 0.06 | | 0.10 | | 0.10 | | 0.07 | |
| | BRSTC | | | | | 0.10 | | 0.07 | | 0.11 | | 0.10 | |
| | | Instruction | | Instruction | | Instruction | | Instruction | | Instruction | | Instruction | |
| | | Spanish | English | Spanish | English | Spanish | English | Spanish | English | Spanish | English | Spanish | English |
| Teacher | OLETC | 0.06 | 0.11 | 0.07 | 0.16 | 0.09 | 0.27 | 0.11 | 0.24 | 0.24 | 0.17 | 0.26 | 0.24 |
| | OLSTC | 0.04 | 0.21 | 0.07 | 0.16 | 0.07 | 0.19 | 0.13 | 0.21 | 0.22 | 0.20 | 0.23 | 0.18 |
| | BRETC | NA | | | | 0.06 | 0.07 | 0.06 | 0.13 | 0.11 | 0.14 | 0.17 | 0.18 |
| | BRSTC | | | | | 0.15 | 0.03 | 0.25 | 0.09 | 0.23 | 0.07 | 0.32 | 0.10 |
| Student | OLETC | 0.85 | 0.79 | 0.82 | 0.73 | 0.85 | 0.67 | 0.83 | 0.69 | 0.71 | 0.77 | 0.70 | 0.70 |
| | OLSTC | 0.87 | 0.74 | 0.82 | 0.76 | 0.90 | 0.79 | 0.83 | 0.77 | 0.71 | 0.75 | 0.70 | 0.78 |
| | BRETC | | | | | 0.82 | 0.67 | 0.94 | 0.68 | 0.88 | 0.82 | 0.87 | 0.88 |
| | BRSTC | | | | | 0.82 | 0.84 | 0.73 | 0.82 | 0.69 | 0.81 | 0.64 | 0.79 |

**Figure 1: Distribution of Standard Scores across six Waves for Oral Language Composite**
English (top panels) and Spanish (bottom panels)
Ignoring Language of Instruction (left panels) and Controlling Language of Instruction (right panels)

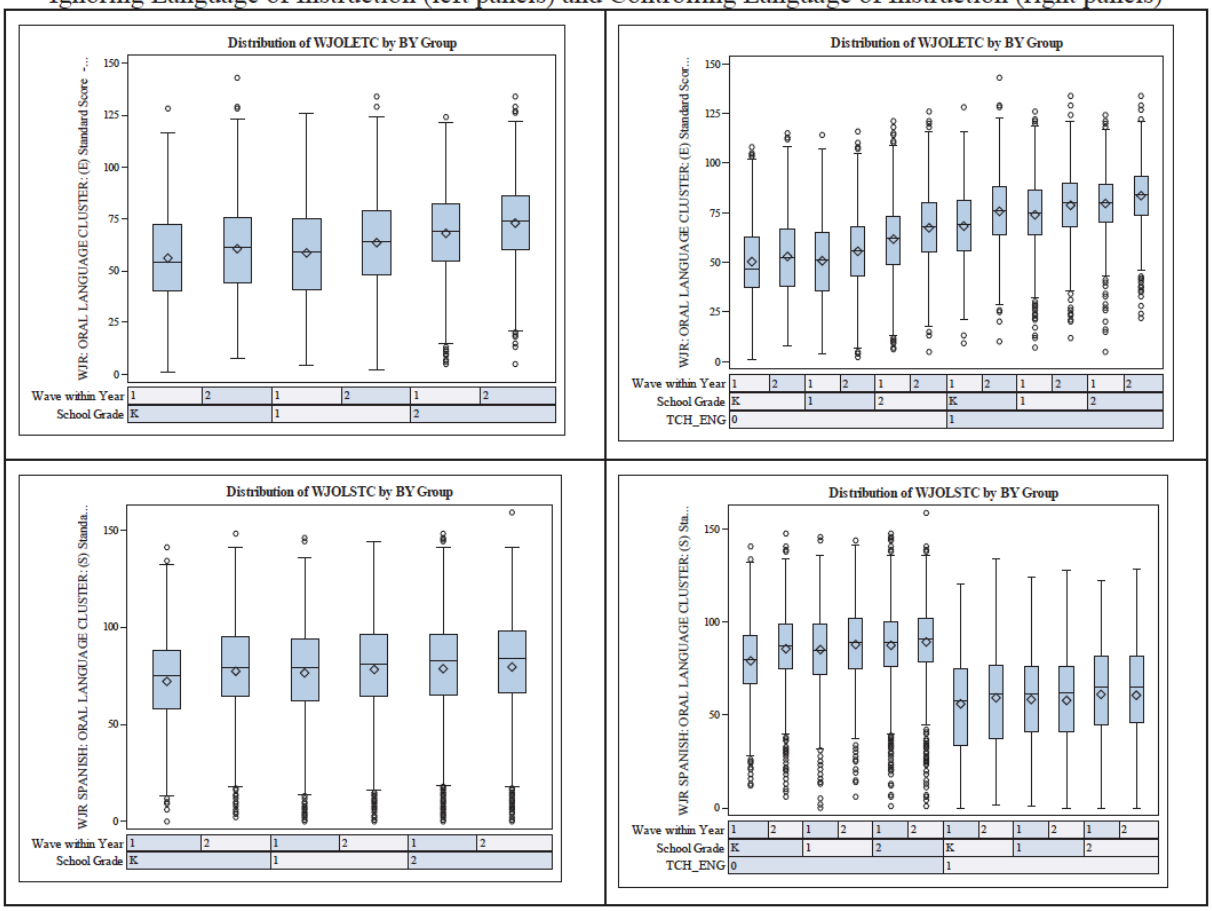

**Figure 2. Mean Plots for Oral Language and Reading by Language of Instruction and Language of Outcome.**