Latent Language Ability Groups in Bilingual Children

Across Three Methods of Assessment

by

Maria Kapantzoglou

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2012 by the
Graduate Supervisory Committee:

Maria Restrepo, Chair
Shelley Gray
Marilyn Thompson

ARIZONA STATE UNIVERSITY

August 2012

ABSTRACT

Differentiating bilingual children with primary language impairment (PLI) from those with typical development in the process of learning a second language has been a challenge. Studies have focused on improving the diagnostic accuracy of language measures for bilinguals. However, researchers are faced with two main challenges when estimating the diagnostic accuracy of new measures: (a) using an a priori diagnosis of children (children with and without PLI), as a reference may introduce error given there is no gold standard for the a priori classification; and (b) classifying children into only two groups may be another source of error given evidence that there may be more than two language ability groups with different strengths and weaknesses or, alternatively, a single group characterized by a continuum of language performance.

The current study tested for the number of distinct language ability groups and their characteristics in predominately Spanish-speaking children in the U.S. without using an a priori classification as a reference. In addition, the study examined to what extent the latent groups differed on each measure, and the stability of language ability groups across three assessment methods in Spanish (standardized tests, language sample analyses, and comprehensive assessment), taking in to account English and non-verbal cognitive skills. The study included 431 bilingual children attending English-only education. Three latent profile analyses were conducted, one for each method of assessment. Results suggested more than two distinct language ability groups in the population with the method of assessment influencing the number and characteristics of the groups.

Specifically, four groups were estimated based on the comprehensive assessment, and three based on standardized assessment or language sample analysis in Spanish. The stability of the groups was high on average, particularly between the comprehensive assessment and the standardized measures. Results indicate that an a priori classification of children into two groups, those with and without PLI, could lead to misclassification, depending on the measures used.

To Gerasimos, our son Dionisis, and my family in Greece

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my mentor, Dr. Adelaida Restrepo, who gave me the opportunity to conduct this study, and assisted me while allowing me room to work in my own way. Her flexibility, openness to my interests, and expertise in a great variety of areas related to bilingualism allowed me to work on topics of my interest, which has been highly motivating, and made my work enjoyable throughout my four years in the doctoral program.

I wish to express my sincere gratitude to my committee members and advisors Dr. Shelly Gray and Dr. Marilyn Thompson, who gave me their time and expertise to better my work. I would like to thank them for their insightful comments, advice, and especially, for their positive feedback and encouragement.

I would also like express my gratitude to Giovanna Moreno, our lab manager, and the students in the lab for their excellent collaboration, and particularly, for the friendly and warm relationship we have developed. Our lab felt like home.

My special thanks to the children and families that participated in the study. I acknowledge the grant funded by IES #R324A080024, without which this study would not have been possible.

Thank you.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Chapter 1

Literature Review

Diagnosis of PLI in Bilinguals

The special characteristics of children growing up bilingually make it difficult to assess their language abilities. For example, bilingual children have diverse language experiences in each language depending on the home, school, and community language input and use. These differences across children lead to differences in growth in each language across children, depending on the quality and quantity of input in each language, and how the two languages influence each other (e.g., Kohnert & Bates, 2002; Restrepo, Castilla, Arboleda, Schwanenflugel, Neuhart Prittchett, & Hamilton, 2010). Therefore, it is difficult to determine what normal language development is for a child at a given age and with a given language experience.

Children with primary language impairment (PLI) are those whose language abilities are significantly below age expectations and do not meet criteria for hearing loss, intellectual disability or autism spectrum disorder (American Psychiatric Association; DSM-5, 2013; Leonard, 1998). Theoretical approaches explain PLI as a linguistic deficit or the result of underlying cognitive and cognitive-linguistic limitations, which affect language, but do not cause intellectual impairment (see Schwartz, 2009 for a review). Language deficits may include poor *grammatical* (Leonard, Eyer, Bedore, & Grela, 1997; Oetting & McDonald, 2001; Owen & Leonard, 2006; Redmond, 2003), *lexical* (e.g., McGregor, Friedman, Reilly, & Newman, 2002; Rescorla, 1989; Thal, O'Hanlon, Clemmons, & Fralin, 1999), and *phonological* skills (e.g., Edwards & Lahey, 1996; Marshall & Van der Lely, 2006; Shriberg, Tomblin, & McSweeney, 1999); whereas, cognitive explanations of PLI refer to limitations in *working memory* (e.g., Archibald & Gathercole, 2006; Dollaghan & Campbell, 1998; Gathercole & Baddeley, 1990; Montgomery, 2000), *speed of processing information* (e.g., Gathercole & Baddeley, 1990; Kail & Leonard, 1986; Montgomery, 2000), *speech perception* (e.g., Burlingame, Sussman, Gillam & Hay, 2005; Tallal & Piercy, 1973, 1974), and *attention and executive functions* (e.g., Hanson & Montgomery, 2002; Norbury, 2005; Weismer, Evans, & Hesketh, 1999).

To identify bilingual children with PLI, development in both languages needs to be considered (e.g., Bedore & Peña, 2008; Bedore, Peña & Garcia, 2005; Restrepo & Kruth, 2000). Low language skills in the first language, for example, do not indicate language impairment when second language abilities are adequate; rather, they may suggest language loss due to low exposure to the first language (e.g., Anderson, 2004; Schiff-Meyers, 1992) or incomplete or protracted acquisition (e.g., Montrul, 2002; Morgan, Restrepo, & Auza, in press; Restrepo et al., 2010). Even when language skills in both languages are low in relation to monolingual children or children with different language experiences, the diagnosis is still not clear because children may be losing skills in the first language while the second language is still developing (e.g., Restrepo & Kruth, 2000; Schiff-Meyers, 1992). Overall, performance of bilingual children on language tasks depends on both language skills and the level of use of each language (e.g., Anderson, 2004; Montrul, 2002; Restrepo & Kruth, 2000).

Further, research suggests that children with PLI and their bilingual peers with typical language development (TLD) have similar linguistic profiles (e.g., Morgan, Restrepo, & Auza, in press; Paradis & Cargo, 2000, 2004; Paradis, Rice, Cargo & Marquis, 2008). For these reasons, language impairment can be misdiagnosed in bilingual populations (e.g., Samson & Lesaux, 2009). Given that bilingual language development is a complex phenomenon, it is important to understand the underlying structure of language ability groups in bilinguals as a first step to better differentiate language ability profiles that could indicate language impairment from those that suggest normal bilingual development.

A number of studies have focused on improving the diagnostic accuracy of language measures for bilinguals (e.g., Burton & Watkins, 2007; Girbau & Schwartz, 2008; Guttiérez-Clellen, Restrepo, Bedore, Peña, & Anderson, 2000; Peña, Iglesias, & Lidz, 2001; Restrepo, 1998; Restrepo et al., 2010); however, researchers are faced with two main challenges when estimating the diagnostic accuracy of new measures: (a) using an a priori diagnosis of language ability (children with and without language impairment) as a reference may introduce error when there is no gold standard for the a priori classification (e.g., Dollaghan, 2004; Kohnert, 2010; Restrepo, 1998); and (b) classifying children into only two groups could be a source of error given evidence that there may be more than two language ability groups with different strengths and weaknesses (e.g., Conti-Ramsden, Crutchley, & Botting, 1997; Rapin & Allen, 1987) or, alternatively, there could be a single group characterized by a continuum of language performance (e.g., Leonard, 1987; 1991; Tomblin & Zhang, 1999).

Identification of language ability groups in the population will help the validation process of experimental measures, resulting in a better criterion against which new measures can be tested. Correct estimation of language ability groups in the population can help develop measures that will better identify these groups. For example, the classification accuracy of a grammatical task can be assessed using a group with grammatical limitations as reference as opposed to aiming at identifying a possibly more heterogeneous group of children with PLI.

Group estimation has to be based on participants' performance on measures (group indicators), which inevitably include measurement error. Developing perfect measures for underlying abilities, such as, language abilities is not a realistic goal; nevertheless, advanced statistical approaches, such as latent model-based analyses, may be used to examine language ability groups and the measures that are best for group differentiation without using an a priori classification as a reference and accounting for measurement error in the group indicators (e.g., Magidson & Vermount, 2002).

A Priori Classification in Bilinguals

The classification accuracy of new language measures is often tested against an a priori classification of the participants, which is based on a variety of sources to triangulate data: standardized norm-referenced tests or other individualized assessments, parent, teacher and/or speech-language pathologist (SLP) questionnaires, and referrals for services (e.g., Ceasar & Kohler, 2007; Kohnert, 2010; Restrepo, 1998); however, the over-representation of children from culturally and linguistically diverse backgrounds in special education

(Artiles & Ortiz, 2002; Artiles, Rueda, Salazar, & Higareda, 2005; Limbos & Geva, 2001; Samson & Lesaux, 2009) has raised a concern with respect to the referral and decision making processes, and the validity of the assessments (e.g., National Research Council, 2002; Restrepo & Silverman, 2001; Skiba, Simmons, Ritter, Kohler, Henderson, & Wu, 2006). Although these issues are not unique to bilingual children, the diverse language and educational experiences in this population add to the complexity of the diagnostic processes at the institutional, professional and test levels.

Referral processes. Error in classification may be introduced at the institutional level when the referral processes for English language learners are not consistent across schools and the efficiency of their implementation also varies (Harry & Klingner, 2006; Skiba et al., 2006). Therefore, using the criteria of a referred sample or a priori classification may lead to error in the criterion. For example, Harry and Klingner (2006) reported great variability on how schools refer and process bilingual children in special education, the adequacy of bilingual assessments, the parents' participation in referral processes, and staff's understanding of the referral processes. Further, based on data from school districts with disproportionate number of minority students, Skiba et al. (2006) found that cultural mismatch between the teachers and students or inadequate training lead to challenges in dealing with behavioral issues, and to increased number of referrals for English language learners. Students that are referred will most likely receive services (e.g., Algozzine, Ysseldyke, & Christenson, 1982; Ysseldyke, Vanderwood, & Shriner, 1997).

Professional judgment. Classification error may also be introduced when the reference groups are formed based on clinical or teacher judgment, professionals' interpretations of interpersonal interactions (e.g., Artiles & Kozleski, 2010; Hammer, Detwiler, Blood, & Qualls, 2004; Roseberry-McKibbin, Brice, & O'Hanlon, 2005). Cultural characteristics play a role in the type and level of participation and communicative interactions in the classrooms, and how these interactions are interpreted (e.g., Duranti, 1997; Skiba et al., 2006). Classroom participation is a complex phenomenon influenced by a large variety of situational and cultural characteristics, and thus it may not be a clear indicator of children's cognitive or language skills (Artiles & Kozleski, 2010). Even when teachers and clinicians use evidence-based instructions or interventions, children's performance related to target skills may be misinterpreted due to sociocultural and psychological factors that may interfere with professional judgment (Artiles & Kozleski, 2010).

Many speech-language pathologists do not have adequate training for working with children from culturally and linguistically diverse backgrounds (e.g., Hammer et al., 2004; Roseberry-McKibbin et al., 2005), leading to inadequate criteria for clinical judgment of PLI in bilingual children. For example, Hammer et al. (2004), based on survey conducted with 213 speech-language pathologists working in public schools, reported that approximately one out of three of them had not received any training for working with children from culturally and linguistically diverse backgrounds. In addition, 18-25% of the participants could not recall whether they had received any related information in one or more courses.

Validity of assessment. The validity of test score interpretations is susceptible to bias when tests are used with children from culturally and linguistically diverse backgrounds (American Educational Research Association, 1999; Cummins, 1988; for reviews, see Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; Solorzano, 2008; van de Vijver & Tenzer, 2004). Bias leads to errors in classification of children with special needs. Potential sources of bias in assessment include cultural mismatch between the skills required for the test and socialization practices (Raykov & Marcoulidis, 2011; Reynolds, 1982; Podsakoff et al., 2003; Solorzano, 2008; van de Van de Vijver & Tenzer, 2004), and use of expressions and translations that are not appropriate for a child's culture and test-taking experience (e.g., Gutierrez-Clellen, 1996; Peña et al., 2001). For example, some words or expressions may be unfamiliar to some children due to dialectal differences, increasing task difficulty (e.g., Abedi, 2004; Peterson, 2000; Podsakoff et al., 2003). Further, some children may not be familiar with procedures such as naming pictures (e.g., Gutierrez-Clellen, 1996), using labels as opposed to descriptions for naming tasks (e.g., Peña et al., 2001), or responding to questions with an obvious response to the listener (e.g., Goldstein, 2004). Such factors, although unrelated to target skills, may influence test results and lead to misdiagnosis of children's language abilities.

Language Assessment Methods in Bilinguals

To reduce bias and improve the diagnostic accuracy in children with PLI in bilingual populations, studies on assessment methods have examined *adaptations of standardized norm-referenced tests* to make tests culturally and linguistically more appropriate (e.g., Burton & Watkins, 2007; Peña et al., 2001; Seiger-Gardner & Brooks, 2008; Windsor, Kohnert, Loxtercamp, & Kan, 2008; Wiig, Secord, & Semel 2006). Studies also examined alternative methods of assessment such as *language sample analysis* (e.g., Gutierrez-Clellen & Hofstetter's, 1994, adaptation to Spanish of Hunt's (1965) procedures; Guttiérez-Clellen et al., 2000; Restrepo, 1998; Restrepo et al., 2010) and some *cognitive/processing tasks* (e.g., phonological working memory: Girbau & Schwartz, 2008; Windsor, Kohnert, Lobitz, & Pham, 2010; speed of processing information: Kohnert, Windsor & Pham 2009) as less culturally biased methods of assessment.

Standardized norm-referenced measures. Clinicians have used standardized norm-referenced measures for diagnostic purposes; however, studies have indicated bias in test score interpretations for bilingual populations (Horton-Ikard & Ellis Weismer, 2007; Peña et al., 2001; Restrepo & Silverman, 2001). In an attempt to minimize cultural and linguistic bias, recent versions of some standardized tests have been adapted for use with bilingual children including normative data for Spanish-English speaking bilingual children (e.g., Preschool Language Scales-4 Spanish, PLS-4 Spanish, Zimmerman, Steiner, & Pond, 2002; Clinical Evaluation of Language Fundamentals-4, Spanish Edition, CELF-4

Spanish, Wiig et al., 2006). For example, the CELF-4 Spanish was modified to include vocabulary used by Spanish-speaking children in the western and southwestern parts of the US; it accepted code switches as a normal bilingual process (Wiig et al., 2006, p. 20), and it was normed with predominately Spanish-speaking or bilingual children in the U.S. Although sensitivity for language impairment reported in the manual is .96 and specificity .85 for the core language measures (Wiig et al., 2006), there has been no study yet to replicate these findings, which is necessary to establish stability in classification accuracy and validity for children living in various bilingual and sociocultural contexts.

Even when a standardized normed-referenced measure is well developed, the classification accuracy depends on the a priori classification accuracy of the participants against which the measure is tested in each study (e.g., Thomas, Lanyon & Millsap, 2009). Different diagnostic criteria for the a priori classification of children into groups with and without PLI, lead to groups with different characteristics across studies; therefore, the accuracy with which a test will identify children with PLI is expected to vary across studies. In bilingual children there are two main factors that increase the probability of misdiagnosis in the priori classification of the participants: (a) the lack of a gold standard for the identification of bilingual PLI that has been replicated across studies and populations, and (b) the high heterogeneity in language experiences and language proficiency in each language (Guttierez-Clellen & Simon-Cereijido, 2007; Kohnert & Bates, 2002; Pearson, Fernández, Lewedeg, & Oller, 1997). Error in the a priori classification used as reference propagates to the error in the measure under development, therefore developing standardized norm-referenced measures with strong validity evidence for identifying two groups, with and without PLI, in bilingual populations is a great challenge.

Language sample analysis. Research indicates that language sample analysis may be a better method of identifying bilingual children from culturally and linguistically diverse backgrounds than standardized tests (e.g., Anderson, 1996; Dunn, Flax, Sliwinski, & Aram, 1996; Guttierez-Clellen & Simon-Cereijido, 2009; Restrepo, 1998). Grammatical errors in utterances and lexical diversity are two of the measures investigated (e.g., Kapantzoglou, Fergadiotis, & Restrepo, 2010; Restrepo, 1998; Restrepo & Gutierrez-Clellen, 2001; Cereijido & Gutierrez-Clellen, 2007).

Grammaticality. Studies suggest that production of grammatical utterances is a good indicator of children's language abilities in Spanish for diagnostic purposes (e.g., Restrepo, 1998; Simon-Cereijido & Gutierrez-Clellen, 2007). For example, Restrepo (1998) conducted a study to identify measures that would differentiate predominately Spanish-speaking children with and without PLI in the U.S. Participants were 62 children 5-7 years old with and without PLI identified a priori based on clinical judgment. Language sample analysis and parent report resulted in 91% sensitivity and 100% specificity, with grammatical errors per Terminal unit (TU) and parent report being the best indicators of PLI. Nevertheless, parent report was biased because parents knew their children's diagnosis, thus, their report was not based only on their own judgment. Given that the a priori classification cannot be perfect, sensitivity and specificity estimates in

Restrepo (1998) may be different with a different sample or method of assessment for the a priori diagnosis of the participants. For instance, Simon-Cereijido and Gutierrez-Clellen, (2007) identified bilingual preschoolers with PLI based on parent concerns, bilingual SLP report and the Bilingual English Spanish Assessment (Gutierrez-Clellen & Simon-Cereijido, 2007; Gutierrez-Clellen, Restrepo & Simon-Cereijido, 2006). In their study sensitivity and specificity for grammaticality were 79% and 100% respectively. Although the classification estimates varied across studies, results suggested that grammaticality separates bilingual groups with and without PLI with high accuracy.

Lexical diversity. Investigators have also examined differences in lexical diversity between children with and without PLI in an attempt to identify measures that indicate children's language abilities based on language samples (e.g., Kapantzoglou et al., 2010; Klee, 1992; Klee, Stokes, Wong, Fletcher & Gavin, 2004; Owen & Leonard, 2002; Thordardottir & Namazi, 2007; Watkins, Kelly, Harbers, & Hollis, 1995). In English, some investigators have used the number of different words (NDW) and/or type-token ratio (TTR) based on a fixed number of words or utterances (e.g., Klee, 1992; Watkins et al., 1995; Thordardottir & Namazi 2007), but these measures are dependent on sample length, which leads to spurious results (Jarvis, 2002; Malvern & Richards, 1997; Tweedie & Baayen, 1998; Vermeer, 2000).

A few studies found significant differences between children with and without PLI in lexical diversity using D (e.g., Klee et al., 2004; Klee et al., 2007; Owen & Leonard, 2002). D (Malvern & Richards, 1997; McKee, Malvern, &

Richards, 2000), which combines an algebraic transformation model and curve fitting to control for sample length problems in lexical diversity measures. Owen and Leonard (2002) examined how language samples from 3-7 year-old children differentiated TLD and PLI groups. They found that younger and older English-speaking children with PLI had lower D than their age-matched peers. Klee et al. (2004) also examined D in spontaneous language samples in 27-68 month old children, and found significant differences in lexical diversity using D in Cantonese-speaking children with and without PLI. Further, Klee et al. ran a discriminant analysis that classified 97.8% (44/45) of the participants correctly using mean length of utterance in morphemes (MLU), age, and D. Later, Klee, Gavin and Stokes (2007) replicated Klee et al.'s (2004) study with British- and American-English speaking children 2;0 to 4;2 years of age classifying 39/47 participants (83%) correctly. In Spanish-English speaking children, Kapantzoglou et al. (2010) found that D and MLU differentiated predominately Spanish-speaking children with and without PLI with 79% accuracy. Together, results indicate that D may be a potential indicator for PLI for Spanish-speaking children when combined with other measures, such as MLU for 5-year-old children or younger.

Cognitive/processing measures. Some authors have argued that processing measures may help identify strengths and weaknesses of children with PLI compared to linguistic measures alone in monolingual (e.g., Campbell, Dollaghan, Needleman, & Janosky, 1997; Laing & Kamhi, 2003) and bilingual children (e.g., Kohnert, Windsor & Pham, 2009; Windsor et al., 2010). Processing measures provide additional information for underlying abilities related to performance on linguistic tasks and are also believed to be less influenced by previous language and cultural experiences than grammatical and semantic language tasks (e.g., Campbell et al., 1997; Rodekohr & Haynes, 2001).

Phonological working memory and speed of information processing are two types of processing skills that have received substantial attention in research in children with and without PLI. Studies suggest that children with PLI appear to have more limited phonological working memory capacity (e.g., Archibald & Gathercole, 2006; Gathercole & Baddeley, 1990; Graf Estes, Evans, & Else-Quest, 2007; Kohnert et al., 2009; Marton & Schwartz, 2003; Thordardottir, Kehayia, Mazer, Lessard, Majnemer, Sutton, Trudeau, & Chilingaryanb, 2011; Weismer et al., 1999; Windsor et al., 2010) and/or poorer processing speed than children with TLD (e.g., Kail, 1994; Kail & Leonard, 1986; Miller, Kail, Leonard & Tomblin, 2001; Thordardottir et al., 2011; Windsor & Hwang, 1999).

Phonological working memory. For English monolinguals, non-word repetition is a phonological working memory measure commonly used for differentiating children with and without PLI for research purposes (e.g., Archibald & Joanisse, 2009; Dollaghan & Campbell, 1998; Graf Estes et al., 2007; Gray, 2003). Non-word repetition has been used in a variety of languages (e.g. French: Thordardottir et al., 2010; Italian: Bartolini, Arfé, Caselli, Degasperi, Deevi, & Leonard, 2006; Spanish: Girbau & Schwartz, 2008; Swedish: Reuterskiold-Wagner, Sahlen, & Nyman, 2005) and in bilingual populations for diagnostic purposes (e.g., Girbau & Schwartz, 2008, Gutiérrez-Clellen & Simon-

Cereijido, 2010; Palladino & Cornoldi, 2004; Windsor et al., 2010). For example, Girbau and Schwartz (2008) found that a nonword repetition task identified PLI in 7-10 year old Spanish-English speaking children with 82% sensitivity and 91% specificity. Windsor et al. (2010) found that bilingual Spanish-English speaking children with and without PLI were classified with 94% sensitivity and 57% specificity. In all studies classification accuracy of nonword repetition is estimated based on how many children of the PLI group score low on the task, but it may be that not all children with PLI have low phonological working memory skills (e.g., Miller et al., 2001). It is possible that there are different subgroups affected in phonological working memory, which if identified correctly, would help determine how accurately non-word repetition measures, for example, identify those groups as opposed to a group such as PLI with possibly diverse profiles (c.f., Catts, Adolf, Hogan, & Ellis Weismer, 2005).

Processing speed. Slow processing speed is another indicator of PLI in English and Spanish-English speaking children (e.g., Kail, 1994; Kail & Leonard, 1986; Kohnert, Windsor, & Ebert, 2009; Kohnert, Windsor & Pham, 2009; Miller et al. 2001; Morgan, Srivastava, Restrepo & Auza, 2009; Windsor & Hwang, 1999). Studies have shown that monolingual children with PLI, as a group, have slower processing speed than peers with TLD, but intragroup variability suggests that only a subgroup of children exhibits slower processing speed (e.g., Kail, 1994; Kail & Leonard, 1986; Miller et al. 2001; Windsor & Hwang, 1999). In this case, a rapid automatic naming task would perform adequately only with that subgroup of children, and it would not be a good measure to classify all children with PLI. The limited number of studies with bilinguals suggests slow processing speed in children with and without PLI as in monolingual populations (e.g., Kohnert, Windsor & Ebert, 2009; Kohnert, Windsor & Pham, 2009; Morgan, Srivastava, Restrepo & Auza, 2009). Morgan et al. (2009) found that a rapid automatic naming (RAN) task differentiated predominately Spanish-speaking children with and without PLI with 65% sensitivity and 71% specificity. Thus, 65% of the children identified a priori as having PLI and 29% of the children identified a priori as with TLD scored below the cut score for PLI based on the RAN task. Within group variability suggested that there were children with different language and processing speed profiles within each of the groups determined a priori as having PLI and TLD. These results, like those in phonological working memory indicate the need to look for different language ability groups of children to improve diagnostic accuracy in bilingual children.

To summarize, some measures, such as standardized tests, grammaticality of sentences in language samples, and non-word repetition, show high classification accuracy of bilingual children with and without PLI; however, there is no gold standard that has been consistently replicated across studies in bilingual populations. Similarly, the studies reviewed so far suggest that more than one measure/task are needed to maximize the diagnostic accuracy of PLI, given a sample that has an a priori classification. One hypothesis is that that dividing participants into only two groups does not reflect the heterogeneity in language abilities in the population (e.g., Bishop, North, & Donlan, 1995; Conti-Ramsden et al., 1997; Rapin & Allen, 1987), thus, measures that address a specific language domain identify only a subgroup of children with respective limitations and not all children with PLI given the diversity in profiles. In addition, error in the a priori classification of the participants, due to the lack of a gold standard for diagnostic purposes in bilinguals, also introduces error when estimating the classification accuracy of new measures. A more refined classification of the sample than a dichotomy (children with and without PLI) would allow test how accurately measures identify a group with respective deficits as opposed to a group with possibly diverse profiles. Latent profile analysis (LPA), a model-based approach, can estimate unobservable groups in a sample based on participants' patterns of performance across a series of measures and indicate to what extent the latent groups differ on each measure without the need of an a priori classification as reference.

The Underlying Structure of Language Ability Groups

Currently data suggests that the classification of children into two groups, with and without PLI (see Figure 1 for a graphical representation of an example), may not be accurate (e.g., Leonard, 1987; 1991; Rapin & Allen, 1987; Tomblin & Zhang, 1999; Van der Lely, 1998, 2005; Wilson & Risucci, 1986), with two alternative hypotheses about the underlying structure of language ability groups in the population: qualitatively different subgroups, in which there are more than two language ability groups that present with different strengths and weaknesses or patterns of performance across language areas (e.g., Conti-Ramsden, 1997; Rapin & Allen, 1987; Van der Lely, 1998, 2005) – see Figure 2 for a graphical representation of an example for this hypothesis; a second alternative is quantitatively different groups in which there are no distinct language ability groups, rather children are characterized by a continuum of language ability, and thus, they differ only in level of language abilities (e.g., Leonard, 1987; 1991; Tomblin & Zhang, 1999) – see Figure 3 for a graphical representation of an example for this hypothesis. If either of these hypotheses is true, dividing children into two groups would introduce error in classification.

One way to tests these competing hypotheses (i.e. two groups vs. more than two groups vs. a continuum of performance) is to employ person-oriented analyses (e.g., Pastor, Barron, Miller, & Davis, 2007). Person-oriented analyses use variances and covariances to directly assess the relationships of the participants as opposed to focusing on the relationships of their scores on the variables (Bauer & Curran, 2004). The goal is to sort participants based on their similarities and differences in patterns of ability levels (e.g., Lubke & Muthén, 2005; Muthén, 2001; Muthén & Muthén, 2000; Pastor et al., 2007).

Some research in monolingual English-speaking children has used person-oriented analyses to examine the number of distinct language ability groups based on children's performance on semantic and grammatical tasks (e.g., Bishop et al., 1994; Conti-Ramsden et al., 1997; Dollaghan, 2004; Tomblin & Zhang, 1999). Results across studies are equivocal in detecting language ability groups that differ in language ability profiles or in the level of language abilities across domains (e.g., Bishop et al., 1994; Conti-Ramsden et al., 1997; Dollaghan, 2004; Tomblin & Zhang, 1999). Some studies have found groups with different language ability profiles, whereas others have found differences in the level of language abilities. I have found no study with bilingual children that examines whether there are more than two ability groups.

Distinct language ability profiles. Some studies using person-oriented analyses have found that not only are there more than two language ability groups, but also that these demonstrate different patterns of language skills affected (i.e., semantic and grammatical; Bishop et al., 1995; Conti-Ramsden et al., 1997). For example, Bishop et al. (1995) examined the diversity of profiles in children with PLI using 90 monozygotic and dizygotic twins with PLI, 7-10 years of age. Four standardized tests were administered, including the Test for Reception of Grammar (TROG; Bishop 1982), the WISC-R Comprehension (Wechsler 1974), a sentence repetition test (Semel et al., 1980), and a word-finding test formed by combining items from a children's and adult's picture-naming scale (McKenna & Warrington 1983; Renfrew, 1991). Concordant twins (monozygotic and dizygotic) formed three main language ability groups including children with global deficits (29%), children with sentence repetition deficits (41%), and children with both sentence repetition and word finding deficits (29%) respectively. Cognitive skills had a positive relationship with language abilities. The percent agreement of the patterns observed by Bishop between concordant twins was 52%, suggesting that “these patterns of language scores are reasonably stable phenomena and they do not merely reflect error of measurement” (Bishop, 1994, p.110).

Conti-Ramsden et al. (1997) found language ability groups with different profiles using different methods of assessment. They studied a sample of 242 7-

year-old children with PLI recruited from 118 language units attached to English mainstream schools. They identified groups based on three different methods of assessment: standardized tests, speech-language pathologist (SLP) or teacher report, and a combination of both. The standardized tests examined, grammar in the receptive modality (TROG; Bishop, 1982), and naming vocabulary, number and word reading skills (the British Ability Scales, BAS; Elliot, 1983), articulation (the Goldman-Fristoe Test of Articulation; Goldman & Fristoe, 1986), narrative skills (The Bus Story; Renfrew, 1991), and general nonverbal abilities (the Raven's Matrices; Raven, 1986). Conti-Ramsden et al. (1997) found six groups based on standardized tests alone: (a) 21% of the children had low receptive grammar (below the 16^{th} percentile) and low normal naming vocabulary abilities (32^{nd} percentile), (b) 6.6% of children had high naming vocabulary and receptive grammatical abilities (above 40^{th} percentile), (c) 12% of children had high naming vocabulary (63^{rd} percentile) and low receptive grammatical and other tested abilities (between 11^{th} and 16^{th} percentile), (d) 9.5% of children had strong receptive grammar (55^{th} percentile) and poorer but above the 16^{th} percentile for the rest of the abilities tested including vocabulary, (e) 34.7% of children had low vocabulary and grammatical abilities (below the 12^{th} percentiles), (f) 10.3% of children had higher on naming vocabulary (52^{nd} percentile) than receptive grammar (30^{th} percentile), remaining within normal range, similarly to group (d). The latter groups differed from group (d) in language abilities other than vocabulary and grammar.

Conti-Ramsden et al. (1997) identified different language ability groups based on the SLP or teacher report and when using the combination of standardized tests and SLP or teacher report than when using standardized tests alone. Teachers or speech-language pathologists provided information with respect to each child's articulation, phonological syntactic and/or morphological, semantic and/or pragmatic abilities. Based on SLP or teacher report they identified the following groups for semantic and grammatical skills: (a) children with low grammatical abilities and good semantic and pragmatic skills; (b) children with semantic and/or pragmatic difficulties but good grammatical skills (c) children with no semantic-pragmatic or grammatical difficulties. Three additional groups extracted did not have different patterns in semantics-pragmatics and grammar than those groups described previous. When standardized tests were considered in combination with SLP or teacher report the following groups were identified: (a) children with good naming vocabulary and poor all other skills tested, with primary syntax/morphology limitations; (b) children with no relevant semantic or grammatical difficulties; (c) children with low language abilities except naming vocabulary; (d) a group with significant variability in grammatical and semantic scores characterized by difficulties in other areas; (e) children who scored poorly on all measures; (f) children with primary semantic and/or pragmatic difficulties and some problems with receptive grammar. As in the Bishop et al. (1995) study, cognitive skills had a positive relationship with language abilities.

Both Conti-Ramsden et al.'s study and Bishop et al.' study examined groups in samples of children identified a priori as having PLI as opposed to using an unclassified sample. When using children diagnosed a priori, the diagnostic criteria may influence the characteristics of the sample and thus, the language ability groups identified. Therefore using an unclassified sample with a modeling based approach, such as LPA, would be a better way to estimate language ability groups in the population (e.g., Collins & Lanza, 2010; Gibson, 1959; Lubke & Muthén, 2007; Magidson & Vermount, 2002; Pastor, Barron, Miller & Davis, 2007).

Differences in level of language ability. Other studies support differences in level of language impairment rather than in the type of language abilities affected (e.g., Dollaghan, 2004; Tomblin & Zhang, 1999). These studies indicate that there is a continuum of performance across all skills – see Figure 3 for a graphical representation of an example for this model. For instance, Tomblin and Zhang (1999) examined groups in a sample of 1933 kindergarteners in Iowa which included children with TLD, PLI and general delay. Children's performance on the receptive and expressive tasks of Test of Oral Language Development – Primary: 2 (TOLD-P:2; Newcomer & Hammil, 1988) were used as indicators of children's grammatical and lexical skills and narrative expression and comprehension based on Cullata, Page, and Ellis (1983). Tomblin and Zhang (1999) extracted six groups that differed mainly in the level of language impairment across grammatical and lexical skills. The groups were characterized by the different levels of scores across grammatical and lexical tasks, with scores ranging from low to high across groups. One group with different pattern scored lower on expressive grammar and expressive narrative subtests; however, this group included only 22 children (1%), and the authors did not provide sufficient information to judge whether this is a real group or outliers scoring differently due to reasons unrelated to their language skills. Tomblin and Zhang concluded that their results support the claim that children with PLI appear to be at the lower end of a continuum (Leonard, 1991) rather than a group with different language ability profile than children with typical language development (e.g., Restrepo, Swisher, Plante, & Vance, 1992).

Dollaghan (2004) examined 620 3- and 4-year old children with and without PLI, using a variety of assessment methods as indicators of PLI, such as standardized tests, language sample analysis, and parent report. For the 3-year-old group, language ability groups were examined based on the mean length of utterance in morphemes (MLU), and the total number of words based on parent report obtained through Language Development Survey (LDS; Rescorla, 1989; Rescorla & Achenbach, 2002). For the 4-year-old group, language ability groups were examined based on the MLU and the Peabody Picture Vocabulary Test–Revised (PPVT-R) scores. For both age groups, results indicated no groups with different profiles. Dollaghan's (2004) results are more difficult to interpret with respect to grammatical skills, given that the only grammatical measure used as an indicator for determining language ability groups was MLU. Further, the use of only two measures restricted the number of patterns that could potentially be identified in person-oriented analyses.

Language ability group estimations are influenced by types of measures used as indicators of semantic and grammatical language abilities, and whether and how the children were identified a priori. For example, Bishop et al. (1995) used the TROG, the WISC-R Comprehension, a test of sentence repetition (Semel et al., 1980), and a word-finding test as group indicators, whereas Dollaghan (2004) used the PPVT-R, LDS and MLU. Further, Conti-Ramsden used the TROG and BAS measures, and Tomblin and Zhang used the TOLD-P:2 measures. The validity of test score interpretations and measurement error are different for the different measures, which in turn impacts the accuracy of the group identification (e.g., Dollaghan, 2004; Tomblin et al., 1997). Estimating language ability groups as latent variables, using structural equation modeling for example, would allow for group estimation accounting for error in measurement for the tasks used as group indicators.

Studies that examine differences in language ability groups when using different methods of assessment with the same sample may help to better understand how the method of assessment influences children's classification based on their language abilities. For example, in the Conti-Ramsden et al. (1997) study different groups were identified in the same sample with different methods of assessment. Group characteristics changed when standardized tests, teacher/SLP report, or a combination of both was used. Additional methods of assessment would be of interest for clinical purposes. For bilingual children, for instance, besides standardized tests, language sample analysis and cognitive/processing measures can be used for diagnosis, although, cognitive/processing measures are used more in research than clinically and they are typically conducted in combination with other tests.

Another significant factor to consider when examining groups is whether the participants were identified a priori as having a particular diagnosis (i.e., PLI), because the characteristics of the sample will vary depending on the diagnostic criteria used, and this will influence the characteristics of the groups identified. For example, Bishop et al. (1995) identified children with PLI based on low scores on any of a set of language measures and substantial discrepancy from non-verbal IQ, a practice no longer deemed valid (e.g., Tomblin et al., 1997). Bishop reported that if criteria for determining PLI were modified, the percentage of concordant monozygotic twins, for example, changed from 54% up to 89%. Conti-Ramsden et al. (1997) classified children as PLI based on their eligibility for special education and need for speech and language treatment in the schools in the UK, which is different across countries and states in the US. Although all children in the sample were identified a priori as having PLI, when language ability groups were estimated, a group of children presented with skills within typical range based on standardized tests. Dollaghan (2004) included children that were not classified a priori as with and without PLI, but she reported how diagnosis changed depending on the language ability measures, which is informative regarding how the type of diagnostic measures may influence the a priori classification of the participants. For instance, depending on whether LDS, number of different words (NDW), MLU or PPVT were considered, for the 3-year-old group, the number of children with typical language ranged from 562-

580 and the number of children with PLI ranged from 36-58; for the 4-year-old group, the number of children with typical language ranged from 554-591 and then number of children with PLI ranged from 32-69. Examining language ability groups in an unclassified sample as some of the previous studies have done (e.g., Dollaghan, 2004; Tomblin & Zhang, 1999) may yield results that represent more accurately the structure of language ability groups in the population.

No studies were found that have examined language ability groups in bilingual populations, although frequent misdiagnosis of their language skills (e.g., Samson & Lesaux, 2009) indicates a need to further examine the current diagnostic criteria including the number and characteristics of language ability groups assumed in the population.

Latent Profile Analysis

Latent profile analysis (LPA) is currently one of the most indicated statistical methods for examining the presence of latent (unobservable) groups in a population (e.g., Collins & Lanza, 2010; Gibson, 1959; Lubke & Muthén, 2008; Magidson & Vermount, 2002; Pastor et al., 2007). The goal of LPA is to classify similar individuals into latent groups based on observed variables, when the number of groups and their sizes are not known a priori through model-based reasoning.

To better understand LPA, a direct analogy can be made to the common factor model. Statistically, both factor analysis and LPA, are latent variable models, thus, they include error free *latent* variables, in other words, unobservable variables that are estimated indirectly through a set of measures (*indicators*).

Also, in both LPA and factor analysis the indicators are observed continuous variables; however, in factor analysis, the latent factor is thought of as a continuous variable; on the other hand, in LPA, the latent factor is categorical (i.e. number of latent groups). In both cases, each model seeks to explain a set of covariances among a set of observed indicators as a function of a latent variable, in such a way that once the influence of the latent variable is taken into account, the observed variables are conditionally independent (Bartholomew, 2007).

The difference in the nature of the latent variable (continuous vs. categorical) carries important implications for the types of phenomena for which each model is more appropriate. In factor analysis, the latent factor often represents a trait that ranges along a continuum (e.g., general intelligence, Spearman, 1904). Performance on the observed variables is influenced by the participant's level on the latent factor: higher scores on the latent variable cause higher performance on the indicators. In LPA, the latent variable takes different values that indicate membership in different groups, such as psychopathy subtypes (e.g., Hicks, Markon, Patrick, Krueger, & Newman, 2004). Each category of the latent variable in LPA is associated probabilistically with a specific pattern of responses on the observed variables. Based on Muthén (2004), factor analysis is a variable-oriented approach, in that it focuses on people's scores on variables to evaluate underlying dimensions and their relationship with the observed scores. On the other hand, according to Collins and Lanza (2010), LPA is a person-oriented analytic method, in that people are the units clustered into groups based on their response patterns. For example, such an approach can help answer questions related to the diagnosis of participants based on their performance on a series of tests.

With respect to other person-oriented techniques, which classify participants into groups, such as logistic regression and discriminant analysis, LPA is advantageous for at least two reasons (Magidson & Vermount, 2002). First, LPA is used to examine latent group membership without an a priori classification. Analyses such as discriminant analysis and logistic regression require an a priori classification of the sample into groups based on a criterion and this classification is used a reference. In LPA, group membership for each participant is estimated based on posterior group-membership probabilities. For each individual's scores across measures, the model estimates the probability of this pattern having been sampled from each group after group parameters have been estimated.

LPA differs from techniques such as traditional cluster analysis in that it is a model-based approach in which the researcher specifies and compares statistically different models that correspond to competing hypotheses (Pastor et al., 2007). As a simple example of model-based reasoning is shown in Figure 4. Variable X represents vocabulary. The dataset lacks information about group membership, but a researcher has reason to believe that the sample may have been drawn from two distinct populations (e.g., children with and without PLI). So, the researcher formulates two mutually exclusive hypotheses: hypothesis A states that the observed data are randomly drawn from a single population, whereas hypothesis B is consistent with the idea that the observed data contributing to the distribution are drawn from two populations. The hypotheses A and B can be expressed in mathematical terms, under the hypothesis A, which suggests that a single normal distribution underlies the data, and under the hypothesis B, which suggests that two normal distributions underlie the data. The hypotheses A and B can be compared through their likelihoods. Then, once the model with the highest likelihood is identified, the researcher can reason back to the observed data and interpret them as having been drawn from a single population, or consisting of a mixture of individuals from two distinct populations.

Specifically, the expectation-maximization (EM) algorithm is often used to search for the maximum likelihood (ML) estimates of the model parameters for each proposed model (Myung, 2003). This is an iterative process which provides results (i.e., model parameters) that maximize the between-group variability and minimize the within group variability as it happens with traditional cluster analysis techniques; in this case the statistical procedure is more rigorous because it involves less subjectivity in determining the best solution (e.g., Madison & Vermount, 2002; Pastor et al., 2007).

In summary, LPA is a rigorous statistical approach with significant advantages for examining latent groups in the population. It is a model-based approach with mathematical machinery that makes possible group specification when heterogeneity is unobserved (i.e. in language ability) and group membership is not known a priori for the participants (e.g., Pastor et al., 2007). Group membership is inferred from the data. Groups are specified based on the interpretability of the results and more objective indicators of model fit. For example, Pastor et al. (2007) contrasted LPA with cluster analysis, discussing the questionable utility or appropriateness of statistics used in cluster analysis and how researchers rely heavily on their own judgment to select the best solution. Therefore, LPA has been chosen as the statistical technique appropriate for examining the questions of this study.

There are no studies that have used LPA to identify language ability groups in children in the population. In prior studies that have examined language ability groups using person-oriented approaches (i) error in measurement of language abilities used as indicators for group estimation may impact the accuracy with which groups are identified; (ii) it is suggested that different methods of assessment (e.g., standardized tests, language sample analysis, teacher or SLP report or a combination of methods) may result in different language ability groups but there is limited evidence for method effects on group estimation within the same sample given limited number of studies and the few types of methods examined; (iii) a priori diagnosis of participants in some studies has error given imperfect measures which may lead to initial misdiagnosis in the sample, and in turn to inaccuracy in group estimation, and (iv) language ability groups have been examined only in monolingual populations.

Purpose of Study

The current study (a) identifies the number and characteristics of latent language ability groups in predominately Spanish-speaking children who speak English as a second language to describe the latent structure of language abilities in this population; (b) specifies groups based on three different methods of assessment in the same sample to examine how assessment method may influence group identification (c) takes advantage of recently developed mathematical algorithms and model-based reasoning using latent profile analysis to identify latent language ability groups in an unclassified sample without using an imperfect a priori classification of the participants as a reference; (d) accounts for the error in measurement of language abilities used as indicators to identify the latent (unobserved) language ability groups with latent profile analysis; and (e) identifies the measures on which these latent groups differ most.

The following questions will be addressed:

1. How many distinct language ability profiles can be identified in 5-to-7-year-old predominately Spanish-speaking children learning English as a second language based on (a) a comprehensive assessment in Spanish including a published norm-referenced standardized test (CELF-4 Spanish; semantic and grammatical tasks), language sample analysis (semantic and grammatical measures), a working memory task, and a speed of processing information task, (b) English language skills as measured by the Structured Photographic Expressive Language Test 3 (SPELT-3; Dawson, Stout, & Eyer, 2003), and (c) non-verbal cognitive abilities as measured by the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2006).

2. To what extent do the latent groups differ on each measure of the comprehensive assessment?

3. How stable are classifications of the participants across three methods of assessment in Spanish (comprehensive, CELF-4 Spanish alone, language sample analysis alone) combined with measures of English and non-verbal skills?

Model-based reasoning through latent profile analysis will be used to examine three hypotheses: that there are two, more than two (3-6) or no distinct language ability groups in the population. It was hypothesized that the method of assessment will influence the number or characteristics of the groups. Further it was hypothesized that grammaticality and phonological working memory measures will yield better language ability group separation than the other measures.

Chapter 2

Method

Participants

Four-hundred thirty one predominantly Spanish-speaking children participated in this study. Participants were selected randomly from a larger study. SES was assessed using children's eligibility for free or reduced-price lunch. All children were recruited from public school and charter programs in Phoenix.

All participants met the following criteria:

1. Children did not have any significant hearing loss based on a pure-tone hearing screening (American Speech and Hearing Association, 1997). Children passed the screening at 500Hz at 25db, and at 1000Hz, 2000Hz and 4000Hz at 20db in both ears.
2. Children were identified as predominantly Spanish speaking if they met the following criteria:
    a. Parents reported the child spoke Spanish more than 50% of the time at home, according to a parent questionnaire.
    b. Teachers reported the child's English language skills were lower than those of a native English speaker at the expressive level, according to a teacher questionnaire.
    c. On a language proficiency scale (Smyk, Restrepo, Gorin, & Kapantzoglou, 2009) that uses story retell as a language elicitation technique, children demonstrated expressive language skills lower than a native speaker level of the same age in English.

d. Children obtained a standard score equal to or lower than 81 on SPELT-3 to exclude balanced bilinguals.

General Procedure

All procedures were approved by the Institutional Review Board at Arizona State University. Each child participated in two sessions of approximately 50 minutes each and in one session of 20 minutes. Children were assessed during the school day in a quiet room in the school.

In the first session, children completed the hearing screening, the story retell task in English, SPELT-3 and the WNV. In the second session, children completed the story retell task in Spanish and CELF-4 Spanish. In the third session, children completed the experimental tasks. Sessions were at least one day apart from each other to avoid participants' fatigue.

Measures

Parent report. All parents filled out a questionnaire requesting demographic information, parents' and child's education history, child's ratings of language skills, child's medical history, family history related to language and learning abilities, and child's exposure to and use of Spanish and English. Parent questionnaires and consent forms were distributed and collected by teachers.

Teacher report. All teachers filled out a questionnaire for each child whose parents agreed to participate in the study. Teachers provided information regarding each child's Spanish and English language abilities, the frequency with which children spoke each language, and concerns regarding children's learning, cognitive, or social skills.

Clinical Evaluation of Language Fundamentals-4, Spanish Edition (CELF-4 Spanish). The CELF-4 Spanish is widely used to assess language skills and to determine eligibility for language services in children, adolescents, and adults 5-21 years old. This study used the Core Language subtests (they identify language disorder). CELF-4 Spanish was normed with predominately Spanish speaking or bilingual Spanish-English populations in the U.S. The standardization sample included 5-7% of children with language impairment. Parental education and occupation varied. Inter-scorer reliability coefficients across subtests that required scoring judgments ranged from .81 to .99, and the standard error of measurement for the Core Language Scores based on split-half reliability coefficients ranged from 2.15 to 3.65 for 5-7 year old children. Sensitivity was .96 and specificity .87 for the core language score at 1 *SD* below the mean.

Wechsler Nonverbal Scale of Ability (WNV). The WNV is used to assess nonverbal cognitive abilities from ages 4-21 years. The norms for the U.S. are based on a standardization sample designed to match the U.S. population on age, sex, race/ethnicity, education level and geographic region. Approximately 88% of the participants in the norming sample spoke English as their first language. Bi/multicultural participants were 60%. Approximately 4.3% of the normative sample included children who were learning English as a second language, children with language disorders, reading and written learning disorders. Inter-rater reliability coefficients ranged from .88 to 1 across all subtests, and the standard error of measurement ranged from 3.67 to 5.41 for full scale score of the short version. The short version of the scale was used for this study. Correlation of the full scale score of the short version with the Wechsler Preschool and Primary Scale of Intelligence – Third Edition (WIPPSI-III; Wechsler, 2002) was .67, with the Wechsler Intelligence Scale for children – Fourth Edition (WISC-IV; Wechsler, 2003) was .58 and with the Wechsler Intelligence Scale for children – Fourth Edition, Spanish (WISC-IV Spanish; Wechsler, 2005) was .67.

The Structured Photographic Expressive Language Test 3 (SPELT-3). The SPELT-3 measures English morphology and syntax skills in children aged 4 to 9 years, 11 months, and it was used as an indicator of children's English language skills in the current study. The standardization sample almost matched the U.S. population with respect to African American and White children, but Hispanic populations and other ethnicities were under-represented. Approximately 7% of the sample was identified with language impairment. There was interrater agreement within one point for 90% of the sample. The standard error of measurement with 95% confidence interval ranged from 2.19 to 2.85.

Language samples and language proficiency. Children completed a story retelling task in English and in Spanish using the books *A Boy, a Dog, and a Frog* (Mayer, 1967) and *Frog on his Own* (Mayer, 1967). Both stories were available in both languages; the two versions were used in random order, one in each language. The two versions were equivalent regarding length, vocabulary, and complexity (Smyk, Restrepo, Morgan, & Kapantzoglou, 2009). Picture support was provided when the test administrator was reading the story and during the child's retell to decrease cognitive load.

English language samples were used to assess each child's language proficiency based on a language proficiency scale, the Spanish-English Language Proficiency Scale (Smyk, Restrepo, Morgan, & Kapantzoglou, 2009), which measures sentence length and complexity, grammaticality, vocabulary, and fluency. Initially, vocabulary, and sentence length and complexity were rated on 1-4 point scale; grammaticality and fluency were rated on a 1-5 point scale. Then, an overall proficiency level was determined ranging from 1 to 5 (1= silent/observer; 2 = a few words or formulaic phrases; 3 = short sentences and phrases with multiple grammatical errors; 4 = full sentences with a few grammatical errors; 5 = native like productions). Language samples were rated immediately after completion of the retelling task by the examiner in each language or at the lab after listening at the audio files. Examiners were fluent Spanish-English or English speaking research assistants who underwent training for using the scale. Only English language samples were used to assess language proficiency in the second language.

Evidence for the validity and reliability of this language proficiency measure has been examined for Spanish-English bilingual children with varying levels of English proficiency. Scale ratings were significantly correlated with the language sample analysis results, for example, for grammaticality with number of grammatical utterances, $r = .73$; vocabulary with number of different words, $r = .62$; and sentence length and complexity with MLU, $r = .66$. Inter-rater reliability for language proficiency levels was 96% (Smyk et al., 2009).

Spanish language samples were used to assess children's language ability in the first language as measured by D and grammatical errors per TU. They were transcribed and coded by the author who is fluent in English and Spanish, and bilingual research assistants with background in linguistics who underwent training. Transcription and coding were completed using the Systematic Analyses of Language Transcripts, research version (Miller & Iglesias, 2008).

Spanish Screener for Language Impairment in Children (SSLIC). The SSLIC (Restrepo, Gorin & Gray, under development) tasks were developed to identify 5-to-7-year-old predominately Spanish-speaking children at risk for PLI in the U.S. The items and tasks were developed using predominantly Spanish-speaking children primarily with low SES. All children were recruited from public school programs in a metropolitan area in the Southwest.

Items and tasks were designed to be sensitive to the linguistic and cultural characteristics to minimize bias. Test development is being conducted in three Phases over four years. The version used for the current study is from the third Phase of development and includes five subscales (Morphology, Sentence Repetition, Antonyms, Spanish Non-word Repetition, and Rapid Automatic Naming). Preliminary analyses conducted for the previous version indicated that average inter-rater reliability estimates for each subscale ranged from .69 to .97. Sentence repetition had inter-rater reliability .69 and coefficients for the other subtests ranged from .91 to .97. Item homogeneity based on Cronbach's alpha ranged from .64 to .92, and exploratory factor analyses suggested unidimensional subtests (Restrepo, Kapantzoglou, Gorin, & Gray, 2011). Psychometric properties in this most recent version used for the current study are expected to be improved.

SSLIC-Spanish non-word repetition. Children repeated recorded non-words presented to them via headphones, for example, *asegerar*. The words followed the Spanish phonotactic rules and were controlled for phonotactic probability and neighborhood density. Each item was scored based on number of phonemes correct. There was a total of 6 items: three 4-syllable words, two 5-syllable words, and one 6-syllable word. The total score on this task was the sum of items correct. All items are administered.

SSLIC-Rapid automatic naming. Children named four familiar items presented eight times each randomly in a table as accurately and as quickly as they could, to assess processing speed. All words were two-syllable words following CVCV syllable structure and they included early acquired sounds to minimize misarticulations. Accuracy was scored based on the sum of errors. Speed was scored in seconds. The final score was the number of errors per second.

Reliability

Inter-rater reliability was estimated for the new SSLIC tasks (i.e. Spanish non-word repetition and rapid automatic naming) and for language sample analysis. For Spanish non-word repetition, two raters scored children's productions from audios independently for 10% of the sample. Interrater reliability was 90% considering 0-1 phoneme difference as agreement. For rapid automatic naming, two raters scored naming accuracy and timed the task independently on site. Interrater reliability for errors per second was 94% considering 0-1 point difference as agreement. For language samples, different raters transcribed and coded 12% of the language samples independently. Inter-rater reliability was estimated 97% for TUs, 86% for grammatical errors, and 93% for NDWs.

Latent profile analysis

LPA was chosen for latent language ability group estimations. LPA is an example of a finite mixture model (McLachlan & Basford, 1988) and its goal is to classify similar individuals or objects into *K* groups based on *p* observed variables, when the number of groups and their sizes are not known a priori. For instance, assume a univariate dataset from 200 individuals. If it is hypothesized that the data are drawn from single population, then they can be described by the normal probability density function (PDF; Gagné, 2006):

$$L_x = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(-0.5)(x_i-\mu)^2}{\sigma^2}}, \text{ where } X|\mu, \sigma^2 \sim N(\mu, \sigma^2) \quad (1)$$

where x_i is a height value for the i^{th} individual, μ is the population mean, σ^2 is the population variance, and L_I is the likelihood value that describes the height of the normal curve for a particular score value. Since the joint probability for a set of independent events is the product of the individual probabilities for each event (e.g., Ross, 2008), the likelihood of obtaining a given sample (under hypothesis A) is the product of the likelihood of obtaining each score individually (Gagné, 2006):

$$L_{Sample|A} = \prod_{i=1}^{N=200} L_x = \prod_{i=1}^{N=200} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(-0.5)(x_i-\mu)^2}{\sigma^2}} \right\} \qquad (2)$$

However, when the data are assumed to represent a mixture of samples from two normally distributed subpopulations, two sets of parameters, one for each subpopulation, need to be specified. Further, the proportion of cases that have been sampled from each population and appear in the data set must also be estimated (Gagné, 2006). To model hypothesis B, two PDF's are required that give the respective likelihoods of the i^{th} datum, given the two sets of population parameters:

$$L_{x|\Theta_1} = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{\frac{(-0.5)(x_i-\mu_1)^2}{\sigma_1^2}}$$, where $X|\mu_1, \sigma_1^2 \sim N(\mu_1, \sigma_1^2)$ (3)

and

$$L_{x|\Theta_2} = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{\frac{(-0.5)(x_i-\mu_2)^2}{\sigma_2^2}}$$, where $X|\mu_1, \sigma_1^2 \sim N(\mu_1, \sigma_1^2)$ (4)

Equations (3) and (4) describes the likelihood of obtaining a particular score x_i from the first population and second subpopulation, respectively.

Assume that φ_1 and φ_2 are the proportions of the cases in the full sample drawn from populations 1 and 2, respectively. Then, overall, the likelihood of obtaining the i^{th} observation is given by the weighted sum of its likelihood in each distribution (Gagné, 2006):

$$L_x = \varphi_1 L_{x|\Theta_2} + \varphi_2 L_{x|\Theta_2} \qquad (5)$$

And similarly to Equation 2, the likelihood of obtaining a specific sample under hypothesis B is given by:

$$L_{Sample|B} = \prod_{i=1}^{N=200} L_x = \prod_{i=1}^{N=200} \left\{ \varphi_1 L_{x|\Theta_2} + \varphi_2 L_{x|\Theta_2} \right\} \quad (6)$$

Based on Gibson (1958) and Lazarsfeld and Henry (1968), the model can be generalized to *K* latent groups and *p* variables. First, Equation 5 can be re-stated as Equation 7:

$$L_x = \sum_{j=1}^{J=2} \varphi_j L_{x|\Theta_j} \quad (7)$$

and then the model can easily generalize to *K* distinct populations:

$$L_x = \varphi_1 L_{x|\theta_1} + \varphi_2 L_{x|\theta_2} + \varphi_3 L_{x|\theta_3} + ... + \varphi_K L_{x|\theta_K} = \sum_{j=1}^{J=K} \varphi_j L_{x|\theta_j} \quad (8)$$

The model can be further extended to the multivariate form to include *p* variables:

$$L_X = \sum_{j=1}^{J} \varphi_j L_{X|\mu_j, \Sigma_j} \quad (9)$$

where, **X** is a vector of observed scores with length p, $\boldsymbol{\mu}_\mathbf{j}$ is a vector of group specific means for the *p* variables, and $\boldsymbol{\Sigma}_\mathbf{j}$ is the group specific variance-covariance matrix of the *p* variables. According to Equation 9, the likelihood of observing a vector of specific scores on a set of *p* variables **X**=[X_1, X_2, …X_p], is equal to the likelihood of observing this set of scores for a person from the first group times the proportion of the first group, plus the likelihood of observing this set of scores for a person from the second group times the proportion of the second group, … , plus the likelihood of observing this set of scores for a person from the j^{th} group times the proportion of the j^{th} group.

Maximum likelihood. Conceptually, the goal of ML estimation is to identify the parameter values that are most probable given the data and the parametric form of the model (Myung, 2003). Specifically, the expectation-maximization (EM) algorithm is often used to search for the maximum likelihood (ML) estimates of the model parameters for each proposed model (Myung, 2003).

In this section, it was assumed that the population parameters in each model (i.e. μ and σ^2 in Equations 1-2; μ_j, σ_j^2, φ_1, and φ_2[1] in Equations 3-6) were known. Typically though, the model parameters have to be estimated from the data. In practice, the estimation procedure is an iterative process during which an algorithm tries out different values for the model parameters (e.g., μ and σ^2) until it identifies the values that are most likely to have generated the data (Brown, 2006; Myung, 2003). To do so, every time the algorithm selects a set of model parameters, the sample likelihood is estimated and recorded. This process is repeated many times, each time using a different set of model parameters. Finally, the set of parameters that yields the highest sample likelihood is chosen as the ML estimates.

For practical reasons, it is more convenient to work with the natural logarithm of the sample likelihoods (Brown, 2006). This does not influence any of the resulting parameter estimates because the log likelihood function is just the likelihood function converted to a more tractable metric. The basic log likelihood function is very simple (Gagné, 2006):

---

[1] When two populations are assumed, $\varphi_2 = 1 - \varphi_1$, and therefore only one of the proportion parameters has to be estimated.

$$\log L_{Sample|A} = \sum_{i=1}^{N=200} \log L_i = \sum_{i=1}^{N=200} \log\left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(-0.5)(x_i-\mu)^2}{\sigma^2}} \right\} \qquad (10)$$

which is very similar to Equation (1) when L_i is a function of the PDF for the univariate normal distribution. To identify the most probable values of the μ and σ^2 under model A, different values of μ and σ^2 are substituted in Equation (10) and the sample log likelihood is estimated for each set of estimates. Then, similarly to working with the sample likelihoods, the set of parameters that corresponds to the highest sample log likelihood are chosen as the ML estimates.

Analysis

Question 1. How many distinct language ability profiles can be identified in 5-to-7-year-old predominately Spanish-speaking children learning English as a second language based on (a) a comprehensive assessment in Spanish including a published norm-referenced standardized test (CELF-4 Spanish; semantic and grammatical tasks), language sample analysis (semantic and grammatical measures), a working memory task, and a speed of processing information task, (b) English language skills as measured by the SPELT-3, and (c) non-verbal cognitive abilities as measured by the WNV.

A LPA was conducted to examine patterns of performance based on ten indicators. Eight out of the ten indicators measured Spanish language abilities. They included the four Core Language subtests of CELF-4 Spanish: Concepts and Following Directions primarily for semantic skills, Word Structure, Recalling Sentences and Formulating Sentences, primarily for grammatical skills; lexical diversity as measured by D and number of grammatical errors per T-Unit (TU;

Gutierrez-Clellen & Hofstetter, 1994; adaptation to Spanish of Hunt's (1965) procedures; Guttiérez-Clellen, Restrepo, Bedore, Peña, & Anderson, 2000; Restrepo, 1998) based on Spanish language sample analysis for semantic and grammatical skills respectively; the SSLIC Non-word Repetition and SSLIC Rapid Automated Naming tasks to assess phonological working memory and speed of processing information. To assist group interpretation the remaining two out of ten indicators included SPELT-3 a measures of English language skills, and WNV a measure of non-verbal cognitive abilities.

LPA was conducted using maximum likelihood estimation within Mplus 6.1 (Muthén & Muthén, 1998–2010). LPA models may be difficult to fit, particularly as the number of groups increases. Specifically, local likelihood maximum may result into invalid parameter estimates. To avoid this, multiple sets of starting values were used (e.g., Collins & Wugalter, 1992). For each estimated model at least 500 sets were run (e.g., Geiser, Lehmann, Corth, & Eid, 2008; Pastor et al., 2007).

Decisions upon the final model were made based on the following fit indices: sample-size adjusted Bayesian Information Criterion (SSA-BIC; Yang, 2006), and the Lo, Mendell and Rubin likelihood ratio test (LMR-LRT; 2001). SSA-BIC may be used to compare two models regardless of the parameterization or the number of latent groups specified. A simulation study by Tofighi and Enders (2007) suggested SSA-BIC and LMR-LRT as the best choices for identifying latent groups. As a first step, the number of latent groups was determined based on SSA-BIC (e.g., Lubke & Muthén, 2005; Marsh, Ludtke, Trautwein, & Morin, 2009; Muthén & Muthén, 2000; Pastor et al., 2007; Tofinghi & Enders, 2007) and LMR-LRT. The smaller the values of SSA-BIC, the better the fit. For LMR-LRT, $p<.05$ suggested statistically significant improvement in fit. Given the lack of a "gold standard" for concluding on a final model, interpretability of the results was also considered in addition to fit indices (e.g., Marsh, Hau & Wen, 2004; Marsh, Hau & Grayson, 2005; Marsh et al., 2009). More descriptive global fit statistics, such as the average group assignment probabilities and entropy (indicator of classification certainty), were taken into account to further understand the classification quality. For a good model, average group assignment probabilities are expected to be above .8 (Rost, 2006), whereas for entropy there is no particular cut-off. The values range between 0 and 1 with larger values suggesting better latent group separation. The prevalence of each group in the final model, and the means and standard errors of scores on each indicator per latent group were estimated.

Question 2. To what extent do the latent groups differ on each measure of the comprehensive assessment?

Ten one-way analyses of variance (ANOVAs) were conducted to examine differences in means between the latent groups estimated in Question 1 on each measure. One-way ANOVAs were conducted using SPSS for Windows Release 11.0.1 (15 Nov 2001). The independent variable, the group factor, included a factor with levels equal to the number of groups identified in Question 1. The dependent variables were the ten measures used as indicators for group estimation in Question 1. The best indicators were determined based on the significance of the pairwise group-mean differences on the measures and the η^2 effect sizes.

Question 3. How stable are classifications of the participants across three methods of assessment in Spanish (comprehensive, CELF-4 Spanish alone, language sample analysis alone) combined with measures of English and non-verbal skills?

Two additional LPAs were conducted as in Question 1 to examine patterns of performance when (i) CELF-4 alone was used for the assessment in Spanish, and SPELT-3 and WNV for English and non-verbal cognitive abilities, and (ii) when language sample analysis alone was used for the assessment in Spanish, and SPELT-3 and WNV for English and non-verbal cognitive abilities. For the first LPA, the following six indicators were used: Concepts and Following Directions from the Core Language subtests of CELF-4 Spanish, indicating primarily semantic skills; Word Structure, Recalling Sentences and Formulating Sentences from the Core Language subtests of CELF-4 Spanish, indicating primarily grammatical skills; SPELT-3 and WNV as measures for English language and non-verbal cognitive abilities. For the second LPA, the following four indicators were used: lexical diversity as measured by D, and number of grammatical errors per TU in Spanish language samples for semantic and grammatical skills respectively; SPELT-3 and WNV as measures for English language and non-verbal cognitive abilities.

Two sets of one-way ANOVAs were conducted as follow-up analyses, as in Question 2, to examine which measures separated the groups best when CELF-

4 Spanish and language sample analysis were used alone respectively as methods of assessment in Spanish. For CELF-4 Spanish as the only method of assessment in Spanish, the independent variable included one factor with levels equal to the number of groups identified with the respective LPA. The dependent variables were the six measures used as indicators for group estimation: Concepts and Following Directions, Word Structure, Recalling Sentences and Formulating Sentences, SPELT-3 and WNV. The best indicators were determined based on the significance of the pairwise group-mean differences on the measures and the η^2. For language sample analysis as the only method of assessment in Spanish, the independent variable included one factor with levels equal to the number of groups identified with the respective LPA. The dependent variables were the four measures used as indicators for group estimation: D, grammatical errors per TU, SPELT-3 and WNV. The best indicators were determined based on the significance of the pairwise group-mean differences on the measures and the η^2 effect sizes.

Results related to language ability groups and group separation based on CELF-4 Spanish as unique method of assessment in Spanish, and language sample analysis as unique method of assessment in Spanish were compared with those from Questions 1 and 2 which examined language ability groups and group separation when a comprehensive assessment was used as a method of assessment in Spanish. Each of these three methods of assessment in Spanish were always combined with SPELT-3 and WNV as measures of English and non-verbal cognitive abilities. First, the stability of groups was analyzed qualitatively by describing changes in the number, interpretation, prevalence, and separation of groups across the three different methods of assessment. Next, data on individual group membership were analyzed. Individuals' most likely group membership was estimated across the three methods and a transition matrix was created to indicate the percentages of individuals that moved from one group to another when different methods of assessment were considered for latent group estimation.

Chapter 3

Results

Preliminary Analyses

The mean age of the participants was 6.52 years (*SD*=.75, range = 5.03 – 8.09). The sample included 50% males and 50% females. All children were from low socio-economic backgrounds based on reduced or free lunch. Descriptive statistics of the major study variables before and after the removal of outliers are presented in Tables 1 and 2.

Data were screened for missing values. The percentage of missing data ranged from 0% to 10.21% across variables. There were a total 115 missing values out of 4310 data points (2.67%). The reasons for missing included inability to test a child at a given time, recording equipment failures, and data entry failures, which are unrelated to the target ability therefore missingness was considered completely at random.

Data were screened for univariate outliers. Outliers were defined as scores that were more than 4 *SD*'s beyond the mean (Kline, 2010; Stevens, 2002). Across all variables, 20 out of 4310 (.46%) of data points were identified as univariate outliers by inspecting frequency distributions of *z* transformed scores. The record forms of each outlier were inspected to explore the reason for which scores deviated significantly from the mean. In most cases, outliers were generated because participants did not participate in a task or did not understand the instructions as indicated by lack correct responses in the trial items. These values were removed and treated as missing data. The patterns of missing data before and after the removal of outliers for each indicator are presented in Table 3.

After the removal of univariate outliers, data were screened for multivariate outliers using the Mahalanobis distance statistic. The Mahalanobis distance statistic is distributed as a χ^2 statistic with degrees of freedom equal to the number of variables. Multivariate outliers were defined as the cases that were associated with *p* values less than .001 (Kline, 2010; Stevens, 2002). A *p* value less than .001 suggests that the null hypothesis that the specific case comes from the same population as the remaining cases is rejected. No multivariate outliers were identified.

After outliers were removed, *z* scores were estimated for all variables/group indicators to minimize the probability of convergence problems and facilitate pattern interpretation. Finally, data were exported to Mplus compatible format for data analysis. A correlation matrix for the major study variables is presented in Table 4.

Main Analyses

Question 1. A LPA was conducted to examine patterns of performance based on a comprehensive assessment. Latent language ability groups were estimated based on children's performance on the following ten measures: the four Core Language subtests of CELF-4 Spanish: Concepts and Following Directions, Word Structure, Recalling Sentences and Formulating Sentences; D and number of grammatical errors per T-Unit estimated from language samples;

the SSLIC Non-word Repetition and SSLIC Rapid Automated Naming tasks; SPELT-3, and WNV.

Models positing between one and five groups were examined considering fit indices (SSA-BIC and LMR-LRT), more descriptive global fit statistics for the classification quality (average group assignment probabilities and entropy), and interpretability of the results. A summary of fit statistics and group characteristics is given in Tables 5 and 6. Figure 5 shows a graphical representation of the latent groups for the estimated solutions that demonstrated statistically significant improvement in fit. SSA-BIC values decreased continuously as the number of groups in the models increased, which suggested improvement in fit as more groups were extracted. However, LMR-LRT indices suggested the difference between the four and five-group models was not statistically significant. Further, one of the classes in the five-group model consisted of only 1.2% of participants. Finally, group interpretability in the more parsimonious solution with four groups was satisfactory. Therefore, the four-group model was selected as best in representing the data.

In the four-group model, SSA-BIC dropped remarkably in relation to its value in the three group model, and LMR-LRT suggested the difference between the two models was statistically significant. Entropy was high (.90) suggesting satisfactory group homogeneity when assuming a four-group solution. Also, the average group assignment probabilities were high, ranging from .93 to .98 across the four groups, which indicates high level of certainty regarding participants' classification. In the four-group model, groups were defined by a combination of differences in profiles and only in level of language performance. Table 7 shows the means and standard deviations of each group on all measures in the four-group model. Group One or the “relatively low group,” with 41.6% of the participants, showed low, but within 1 *SD* from the mean, language ability on all language domains: grammatical, semantic, working memory and speed of processing information in Spanish, English and cognitive non-verbal. Group Two, or the “relatively-low and slow group,” with 7.3% of participants, presented a very similar profile to the “relatively low group,” with the exception that its speed of processing information abilities as measured by RAN were remarkably lower. Group Three or the “average group,” with 37.9% of the participants, performed near average on all measures. Finally, Group Four or the “high group,” with 13.4% of participants, had overall similar patterns of performance as the “average group” but showed greater language abilities in all domains except grammar as measured by GETU and speed of processing information as measured by RAN.

Question 2. Ten one-way ANOVAs were conducted to further examine effect sizes and statistical significance for differences between the latent groups on each of the measures in the comprehensive language assessment. The independent variable was the group factor with four levels. The dependent variables were the ten measures used as indicators for group estimation in Question 1: Concepts and Following Directions, Word Structure, Recalling Sentences and Formulating Sentences, D, grammatical errors per TU, SSLIC Spanish non-word repetition, SSLIC Rapid Automatic Naming, SPELT-3, and WNV. Given the large number of ANOVAs, statistics are presented in Table 8.

Based on the estimated effect sizes, η^2, there was a strong relationship between groups and the following five measures: RAN, Formulating Sentences, Concepts and Following Directions, Word Structure and Recalling Sentences. Effect sizes across these measures ranged from .74 to .52, respectively. There was a moderate relationship between groups and the SPELT-3, SNWR, and WNV, with effect sizes of .26, .24, and .23, respectively. The weakest relationship was between groups and the language sample measures, D and GETU, with effect sizes of .12 and .08 respectively.

Follow-up tests were conducted to evaluate pairwise differences among the group means and identify whether particular pairs of latent groups differed significantly on each measure. Post hoc comparisons were conducted with the Modified Shaffer Sequential Procedure (Shaffer, 1986). In Figure 6, pairwise comparisons that were not statistically significant have been marked with a circle. Given that *z*-score means have been used, the magnitude of the differences between groups is similar to Cohen's *d*. There were statistically significant differences in group means for the great majority of the comparisons. Mean differences between the "relatively low group" and the "relatively-low and slow group" were not statistically significant for any of the measures except RAN. Mean differences between the "relatively-low and slow group" and the "average group" were statistically significant for all measures except D. Mean differences between the "average group" and the "high group" were statistically significant for all measures except GETU. Mean differences between the "relatively low group" and the "average group" or the "high group" were statistically significant for all measures except RAN. RAN differentiated only between the “relatively low and slow group” and the remaining groups.

Question 3. Two LPAs were conducted to examine patterns of performance based on two different methods of assessment in Spanish--standardized normed-referenced test (CELF-4 Spanish) alone and language sample analysis alone--taking into account English and non-verbal abilities in both cases. Specifically, for the first LPA, latent language ability groups were estimated based on children’s performance on the following six measures: the four Core Language subtests of CELF-4 Spanish: Concepts and Following Directions, Word Structure, Recalling Sentences and Formulating Sentences; SPELT-3, and WNV.

Models positing between one and four groups were examined considering fit indices (SSA-BIC and LMR-LRT; 2001), more descriptive global fit statistics for the classification quality (average group assignment probabilities and entropy) and interpretability of the results. A summary of fit statistics and group characteristics is given in Tables 9 and 10. Figure 7 shows a graphical representation of the latent groups for the estimated solutions that demonstrated statistically significant improvement in fit. SSA-BIC values decreased continuously as the number of groups in the models increased, which suggested improvement in fit as more groups were extracted. However, LMR-LRT suggested the difference between the four-group and three-group models was not statistically significant. Also, group interpretability in the more parsimonious solution with three groups was satisfactory. Therefore, the three-group model was selected as best in representing the data.

In the three-group model, SSA-BIC dropped markedly in relation to its value in the two group model, and LMR-LRT suggested the difference between the two models was statistically significant. Entropy was high (.88) suggesting satisfactory group homogeneity when assuming a three-group solution. Also, the average group assignment probabilities were high, ranging from .92 to .96 across the three groups, which indicates high level of certainty regarding participants' classification. In the three-group model, groups are defined only by quantitative differences. Table 11 shows the means and standard deviations of each group on all measures. Group One or the "relatively low group," with 47% of the participants, showed low grammatical and semantic skills as measured by CELF-4 in Spanish, but within 1 *SD* from the mean, and low English and non-verbal abilities as measured by SPELT-3 and WNV. Group Two or the "average group," including 39% of participants, showed average language abilities across domains. Finally, a Group Three or the "high group," with 14% of participants, showed high language ability skills across domains.

Six one-way ANOVAs were conducted to further examine effect sizes and statistical significance for differences between the latent groups on each of the measures based on CELF-4 Spanish scores, as a method of assessment in Spanish, while taking into account English and non-verbal cognitive skills. The independent variable was the group factor with three levels. The dependent variables were the six measures used as indicators for group estimation: Concepts and Following Directions, Word Structure, Recalling Sentences and Formulating Sentences, SPELT-3, and WNV. Results are presented in Table 12. Based on the estimated effect sizes, η^2, there was a strong relationship between groups and the following measures: Formulating Sentences, Concepts and Following Directions, Word Structure and Recalling Sentences. Effect sizes for these measures ranged from .73 to .54 respectively. There was a moderate relationship between groups and the SPELT-3 and WNV with effect sizes .27 and .24, respectively.

Follow-up tests were conducted to evaluate pairwise differences among the group means and identify whether particular pairs of latent groups differed significantly on each measure. Post hoc comparisons were conducted with the Modified Shaffer Sequential Procedure (Shaffer, 1986). All pairwise comparisons were statistically significant among all groups across all measures.

A second LPA analysis was conducted to examine patterns of performance based on language sample analysis alone as a method of assessment in Spanish, taking into account English and non-verbal skills. Specifically, latent language ability groups were estimated based on children's performance on the following four measures: D, GETU, SPELT-3, and WNV.

Models positing between one and four groups were examined considering fit indices (SSA-BIC and LMR-LRT; 2001), interpretability of the results and more descriptive global fit statistics for the classification quality (average group assignment probabilities and entropy). A summary of fit statistics and group characteristics is given in Tables 13 and 14, respectively. Figure 8 shows a graphical representation of the latent groups for the estimated solutions that demonstrated statistically significant improvement in fit. SSA-BIC values decreased continuously as the number of groups in the models increased, which suggested improvement in fit as more groups were extracted. However, LMR-LRT suggested the difference between the four-group and three-group models was not statistically significant. Also, group interpretability in the more parsimonious solution with three groups was satisfactory. Therefore, the three-group model was selected as best representing the data.

In the three-group model, SSA-BIC dropped remarkably in relation to its value in the two group model, and LMR-LRT suggested the difference between the two models was statistically significant. Entropy was .72, suggesting moderate group homogeneity when assuming a three-group solution. Also, the average group assignment probabilities were moderate to high, ranging from .87 to .92 across the three groups, which indicates satisfactory (above 80%) level of certainty regarding participants' classification. In the three-group model, groups are defined different language ability profiles. Table 15 shows the means and standard deviations of each group on all measures. Group One or the "low-grammar group," with 6% of the participants, showed low grammatical skills as measured by GETU, and average semantic English and non-verbal abilities as measured by SPELT-3 and WNV. Group Two or the "average-Spanish and relatively low-English group," including 50% of participants, showed average semantic and grammatical skills and lower, but within 1 *SD* from the mean, English and non-verbal abilities. Finally, Group Three or the "average-Spanish and high-English group," with 44% of the participants demonstrated average semantic and grammatical skills, but higher English and non-verbal abilities.

Four one-way ANOVAs were conducted to further examine effect sizes and statistical significance for differences between the latent groups on each of the measures based on language sample analyses as a method of assessment in Spanish, taking into account English and non-verbal cognitive skills. The independent variable was the group factor with three levels. The dependent variables were the four measures used as indicators for group estimation: D, grammatical errors per TU, SPELT-3 and WNV. Results are presented in Table 16. Based on the estimated effect sizes, η^2, there was a strong relationship between groups and the following two measures: grammatical errors per TU and SPELT-3, with effect sizes of .60 and .50 respectively. There was a moderate relationship between groups and WNV ($\eta^2 = .30$), and a weak relationship between groups and D ($\eta^2 = .07$).

Follow-up tests were conducted to evaluate pairwise differences among the group means and identify whether particular pairs of latent groups differed significantly on each measure. Post hoc comparisons were conducted with the Modified Shaffer Sequential Procedure (Shaffer, 1986). All pairwise comparisons were statistically significant among all groups across all measures.

Table 17 provides the stability of classification of the participants across the three different methods of assessment in Spanish addressed in the current study: comprehensive assessment (standardized norm-referenced measure, language sample analysis, processing/cognitive measures), standardized norm-referenced measure alone, language sample analysis alone; measures are always examined in combination with English and non-verbal abilities measures. Stability in classification was ranged from moderate to high with 94% - 99% of the participants remaining in the same or similar group when comprehensive assessment or CELF-4 Spanish alone was used in Spanish; 93%-97% remaining in the same group when comprehensive assessment or language sample analysis was used; and 82%-95% of participants remaining in the same or similar group when language sample analysis or CELF-4 Spanish were used. So, participants' classification stability was somewhat higher between comprehensive assessment and CELF-4 Spanish alone than between comprehensive assessment and language sample analysis alone.

Chapter 4

Discussion

The main purpose of the current study was to determine the number of latent groups and their characteristics in bilingual, predominately Spanish-speaking children living in the U.S. based on a comprehensive assessment in Spanish, which included (a) a published standardized norm-referenced test (CELF-4 Spanish); language sample analyses (D and GETU); working memory (SNWR) and speed of processing information (RAN) tasks; (b) English skills as measured by SPELT-3, and (c) non-verbal cognitive skills as measured by WNV. Further, this study examined to what extent the latent groups differ on each measure, and how different methods of assessment in Spanish (comprehensive assessment, standardized norm-referenced test alone and language sample analysis alone) may influence the language ability group classification.

The current study is the first to examine the presence of latent language ability groups in bilingual and in Spanish speakers in a bilingual context; although previous studies have examined language ability groups in monolingual English-speaking children. Results indicated that there were four latent groups based on the comprehensive assessment, and three based on standardized assessment or language sample analysis in Spanish. Two groups with different language ability profiles, one estimated with the comprehensive assessment and one with language samples analysis in Spanish, had proportions similar to the prevalence of PLI in the population. The stability of participant classification was high on average (above 82%), particularly between the comprehensive assessment and the standardized measures.

The current study investigated language ability groups and the measures that separated those best when no a priori classification was used as reference. The study used LPA, a statistical approach that uses model-based reasoning and maximum likelihood estimation to infer group membership for each participant from the data. Also, given the latent nature of the groups, the models accounted for measurement error in each indicator used for group estimation. LPA is currently one of the most accurate statistical methods for examining the presence of latent groups in a population (e.g., Collins and Lanza, 2010; Gibson, 1959; Lubke & Muthén, 2007; Magidson and Vermount, 2002; Pastor, et al., 2007). No previous studies to my knowledge have used LPA to investigate language ability groups in monolingual English speaking or bilingual children.

Latent Group Structure Based on a Comprehensive Assessment

The LPA suggested four latent language ability groups based on a comprehensive assessment: a "relatively low group," with low language abilities across abilities tested, but on average within 1 *SD* from the mean; a "relatively-low and slow group," with similar pattern of performance to the "relatively low group" but remarkably lower processing speed abilities and error rates as measured by RAN; an "average group," with average language abilities across abilities tested; and a "high group" with the highest language abilities. Thus, results were supportive of the hypothesis of more than two distinct language ability groups in the population, and indicated one group with a distinct language ability profile, and three groups that differed only in level of language abilities, when using the comprehensive assessment.

Number and type of groups. Finding more than two distinct language ability groups rather than a single group (continuum of language performance) or two distinct groups (with and without PLI), as assumed in current clinical practice for diagnostic purposes, is consistent with Conti-Ramsden et al.'s (1997) and Bishop et al.'s (1995) studies on English monolingual children, which also used person-oriented statistical approaches for group identification. All three studies identified more than two distinct groups based on semantic and grammatical skills, despite differences in their methodology and sample characteristics, which strengthens the findings. Specifically, studies used different semantic and grammatical measures: all three studies used standardized norm-referenced tests, while Conti-Ramsden, also used SLP or teacher report; the current study used language samples analysis, working memory and processing speed tasks in addition to the CELF-4 Spanish, the standardized test. Also, the age and language of the participants varied across studies: Bishop et al. included 7-10 year old English-speaking children, Conti-Ramsden et al. included 7-year-old English-speaking children, and the current study included 5-7 year old predominately Spanish-speaking children in the U.S.

Besides measures, age-ranges and language differences across studies, studies also differed on the a priori diagnosis of participants. Bishop et al. and Conti-Ramsden et al. identified participants a priori as having PLI, whereas the current study included an unclassified sample with and without PLI. The current study identified four groups using a comprehensive assessment, Conti-Ramsden et al.'s study identified six, and the Bishop et al.'s study found three main groups.

Given that the three studies used different indicators, groups cannot be compared directly. Nevertheless, all studies included semantic and grammatical measures as group indicators, and thus, some comparison can be made at the construct level. The current study identified children who scored low across grammatical and semantic measures, which is consistent Bishop et al. (1995) and Conti-Ramsden et al. (1997) studies, who also found a group of low grammar and low semantic group. Further, Bishop et al. (1995) also identified groups with primary limitations in sentence repetition, but not in word finding, although the reverse pattern was not observed. Conti-Ramsden et al. (1997) found groups with low grammatical and high semantic skills, and vice versa. In contrast, in the current study, grammatical and semantic measures differentiated the latent language ability groups primarily due to differences only in the level of ability. Children did not present with remarkable differences in profiles in semantic and grammatical skills with the comprehensive assessment. Different number and type of group indicators may influence the number and type of groups extracted as suggested by the results across the different methods in current study. In the present study, the number of groups extracted with the same sample was reduced from four to three when only CELF-4 Spanish or language sample analyses were considered in Spanish.

The RAN measure results in the low and slow group, which suggests that some children with low language ability demonstrate slow processing speed with high error rates (e.g., Kail, 1994; Miller et al., 2001; Morgan, Srivastava, Restrepo & Auza, 2009). This finding is not surprising based on previous studies examining processing speed in children with PLI. For example, using data from five different experiments, Kail (1994) found that children with PLI had three times greater response times than children without PLI; however, not all children with PLI demonstrated that deficit performance speed, which has led to inconsistency in the estimation of classification accuracy of similar tasks (e.g., Kail, 1994; Miller et al., 2001; Morgan et al., 2009). The main difference between this and previous studies is that previous studies use an a priori classification to determine the groups in their sample (e.g., Morgan et al., 2009; Kohnert et al., 2009), whereas in the current study groups were identified without an a priori classification. When participants are diagnosed a priori, depending on the diagnostic criteria, sample characteristics may vary. The current study with an unclassified sample indicates that RAN may help identify a language ability group with a different profile, slower processing information skills. These results explain why slow processing skills measures are not good in identifying PLI in general, but they may be good in identifying a subgroup of children with low language ability skills. Considering that the group portion (7%) is similar to the prevalence of PLI, it may be that this group includes children with PLI, although some studies suggest that not all children with PLI have slower processing speed (e.g., Miller et al., 2001).

Finding groups with different language ability profiles contrasts to results from Dollaghan (2004) and Tomblin and Zhang (1999) studies on monolingual

English-speaking children. Two possible explanations can account for differences from the Dollaghan study: age differences between studies and the measures used. Dollaghan examined younger children than studies that found groups with different profiles. Dollaghan examined 3- to 4-year-old children whereas, Bishop et al. (1995) and Conti-Ramsden et al. (1997) and the current study examined school-age children. Perhaps, language ability profiles appear different at younger ages, especially when grammatical skills are developing, which are highly correlated with vocabulary development (Castilla et al., 2009 for English as second language; Kohnert, Kan, & Conboy, 2010). Further, the Dollaghan study used only two measures per age group, total number of words and MLU for the 3-year-old group, and PPVT-R and MLU for the 4-year-old group, which may have reduced the number of groups identified. Differences in the languages spoken, although could be influencing group characteristics to some degree, should not determine whether the structure of language abilities in the population is characterized by a continuum of performance from low to high levels or distinct groups with different profiles.

The group estimation technique may also account for differences in the groups extracted. For example, Tomblin and Zhang (1999) used semantic and grammatical measures as did the current study, and cluster analysis, a person-oriented approach for group estimation. Tomblin and Zhang found six groups that differed primarily in level in semantic and grammar skills, and given the relatively large number of groups extracted results suggested a continuum of performance from low to high ability levels using the TOLD-P:2 as the authors discussed. In the present study, when CELF-4 Spanish was used alone for assessment in Spanish, results also indicated only quantitative differences between groups.

Differences between the current study and the Tomblin and Zhang study (1999) in the number of groups extracted may be related to the type of analysis used. Tomblin and Zhang used cluster analysis to identify groups; however, they did not provide sufficient information on how decisions were made on the final solution. In the current study, improvement in model fit was not statistically significant when more than three groups were extracted. Nevertheless, in both studies, groups differed in the level of language ability as indicated with a set of subtests from a standardized normed-referenced test. Given the use of a single battery to assess semantic and grammatical skills, method effects could be related to the high correlations between the two domains.

Group Stability across Different Methods of Assessment

Number and type of groups. The number and type of language ability groups identified differed when different and shorter methods in Spanish were considered. When only a subtests from the standardized normed-referenced test were entered (CELF-4 Spanish), the LPA suggested three groups: a "relatively low group," with low Spanish abilities across abilities tested but on average within 1 *SD* from the mean; an "average group" with average language abilities across abilities tested; and a "high group" with the highest language abilities.. Tomblin and Zhang (1999) discussed that "most clinical measures are not constructed around linguistic models that make principled claims concerning separate linguistic modules." Perhaps, this could partially explain the lack of difference in performance between semantic and grammatical tasks on CELF-4 Spanish. Interestingly, with this method, the lower group performs near the sample average for all measures and includes a large portion of the participants (47%). The CELF-4 Spanish does not identify a group resembling the 7-10% rate of PLI in the monolingual population (Tomblin et al., 1997).

When language sample analysis was used in Spanish with SPELT-3 for English and WNV for non-verbal cognitive skills, LPA results also suggested three groups: a "low-grammar group," with low Spanish grammatical skills, with average Spanish semantic, English and non-verbal skills; an "average Spanish and relatively low-English group," with average language abilities in Spanish, and relatively low English and non-verbal skills, but on average within 1 *SD* from the mean; and an "average Spanish and high-English group," with average language abilities in Spanish and relatively high English and non-verbal skills. Thus, when using only language sample analysis for the assessment in Spanish, results were supportive of the hypothesis of more than two distinct language ability groups in the population and indicated three groups with different language profiles. In this case, the rate of PLI in the monolingual population seems to be also similar to the prevalence of PLI in the monolingual population (Tomblin et al., 1997).

Semantic and grammatical measures did not reveal groups with different profiles when the comprehensive assessment or only the CELF-4 Spanish were used as methods in Spanish, but they did so when only Spanish language samples were used. When the comprehensive assessment or the CELF-4 Spanish alone were used, children who scored low on semantic tasks also scored low on grammatical tasks, and children who scored high on semantic tasks also scored high on grammatical tasks. However, when only Spanish language sample analysis was used for Spanish assessment, one of the groups (6% of the sample) presented with selective grammatical difficulties. The latter group scored on average 2 *SD* above the mean on grammatical errors per TU (high scores indicate low ability) and .5 SD below the mean on D. This is consistent with one of the groups Conti-Ramsden et al. (1997) identified, in which 12% of the participants performed in the 63^{rd} percentile on naming vocabulary tasks, and in between the 11^{th} and 16^{th} percentile on receptive grammar tasks. As in Conti-Ramsden et al. (1997), children's classification is influenced by the method of assessment and the specific tasks used to assess a particular type of language abilities, which Tomblin and Zhang (1999) also reported. Also, the current study included only expressive tasks as opposed to Conti-Ramsden et al. (1997) who assessed grammar through receptive measures. It may be that differences in language modality also influenced the results. Nevertheless, despite differences in the portion of low-grammar groups, in both studies the percentage of children with grammatical limitations is near the range of prevalence of PLI in monolingual populations.

Low language ability groups. The "low groups" identified across methods should be interpreted with caution because their proportions are large: 42% with the comprehensive assessment in Spanish, 47% with the CELF-4 Spanish alone, and 50% with the language sample analysis alone (average Spanish – relatively low-English group). Participants showed low performance on both Spanish and English relative to the other groups; however, regardless of the method of assessment in Spanish, “the low groups” scored within one standard deviation from the mean across measures and their English abilities on average appeared near the mean. Such results suggest that the relatively low performance does not necessarily indicate PLI. Nevertheless, there are two groups in the solutions across the three methods with proportions similar to the PLI prevalence and different profiles: the “relatively-low and slow group” based on comprehensive assessment and the “low-grammar group” based on language sample analysis. Processing speed and grammatical abilities were on average 2 *SD* from the mean. Also, both groups demonstrated low English skills consistently with the expected profile of children with PLI. Perhaps grammatical and processing speed measures are more sensitive to PLI.

Various factors may be related to low performance on language tasks in bilingual children, including socio-economic status, and incomplete or protracted acquisition and language loss (e.g., Anderson, 2004; Montrul, 2002; Morgan, Restrepo, & Auza, in press). For example, children in the current study are primarily from low socio-economic backgrounds, which is frequently associated with low test scores and low vocabulary levels, which impact the range of semantic skills (e.g., Campbell, Bell & Keith, 2001; Horton-Ikard & Weismer, 2007; Restrepo et al., 2006). In addition, these children attend English-only schools, and thus Spanish, their first language, slows in development especially in grammar (Restrepo, et al, 2010) and semantic skills. In bilingual language development, low scores in both languages may result from language loss or protracted acquisition of the first language due to low exposure to the native language in some contexts, while the second language is still developing (e.g., Anderson, 2004; Morgan et al., in press; Montrul, 2002; Restrepo & Kruth, 2000; Schiff-Meyers, 1992), or due to the unique linguistic characteristics of the bilingual language system, which is different from that of monolinguals (e.g., Volterra & Taeschner, 1978; Genesee, 1989; Grosjean, 1989).

Processing measures of working memory and processing speed in addition to language measures were selected because studies suggest that such measures may be less influenced by language experience and are useful for differentiating children with and without PLI (e.g., Campbell et al., 1997; Kohnert et al., 2009; Rodekohr & Haynes, 2001; Windsor et al., 2010). However, performance on these measures may also be related to language experiences in each language (e.g., Windsor et al., 2010; Storkel, 2001; Vitevitch & Stamer, 2006). For example, Windsor et al.'s (2010) study indicated different performance on SNWR between monolingual and bilingual children. The level of familiarity with the phonemes or the phoneme combinations in SNWR may vary depending on the amount of exposure to Spanish, which may influence the ability to recall the words correctly (Storkel, 2001). The results in this study suggest that SNWR performs like the other language measures, unlike the processing speed measure.

Measures that assess potential for learning, such as dynamic assessment or longitudinal data, used in combination with the current measures could validate children's language abilities. For example, Lei et al. (2010) investigating early predictors of reading skills in Chinese children from 3 to 8 years old, in a 6-year longitudinal study, used growth mixture modeling, and differentiated a group with language difficulties which caught up and developed normal literacy skills, from a group with language difficulties which also showed literacy deficits at age 8. Both groups had showed difficulties in morphology and rapid automatic naming, which is in agreement with the characteristics of the “relatively-low and slow group” and the “low-grammar group” in the current study. Also, dynamic assessment may facilitate diagnosis of low performing participants (e.g., Anderson, 2001; Roseberry & Connell, 1991; Peña et al., 2001; Kapantzoglou, Restrepo & Thompson, 2012). For example, Kapantzoglou et al.’s (2012) study suggested that dynamic assessment of word-learning skills is a potentially good indicator of language abilities in predominately Spanish-speaking children. Anderson (2001) found that dynamic assessment of children’s grammatical-rule learning abilities may assist in diagnosing PLI in Spanish-speaking children. Focusing more on the ability to learn rather than on current language skills that are influenced by experiences, dynamic assessment may be less vulnerable differences in previous language experiences in bilingual children with low SES and low parental education, such as in the current study.

Stability of participant classification. The participant classification stability across methods can provide additional information on the group consistency. Overall, stability was high with participants remaining in groups with some relatively low ability from one method to another. Participants’ classification was more stable between the comprehensive assessment and the CELF-4 alone, with 98% of participants belonging in “the relatively low groups” regardless of method of assessment. A lower percentage (76%) of participants moved from the “relatively-low and slow group” based on comprehensive assessment to the “the relatively low group” based on CELF-4 assessment.

When comprehensive assessment is compared to language sample analysis for Spanish, stability in classification is high if both Spanish and English are considered. However, if only Spanish is taken into account, the stability in classification across methods drops remarkably. For example, 77% of children in the “relatively low group” based on comprehensive assessment moved to the “average-Spanish and relatively low-English group” based on language sample analysis, but only 6% of the children move to the “low-grammar group”. Adding up the two groups with low abilities in one of the two languages based on language sample analysis (77% + 6%), the percentage of participants that remains in groups with some relatively low ability from comprehensive assessment to language sample analysis is high.

The participants in the “relatively-low and slow group” and in the “low grammar group,” who could potentially present with PLI, do not coincide. Only 10% (3) of the children with slow processing speed presented with low grammar abilities as well. In addition, children with slow processing speed performed within 1 *SD* from the mean on all other measures. Thus, if this group does include children with PLI, they would not be identified using a battery that assesses only grammatical and semantic abilities. Further analyses could examine whether two groups with different profiles occur (one with low grammar and one with low processing skills) occur if groups are estimated based on language sample analysis and processing tasks in Spanish, and considering English and non-verbal cognitive skills.

Group Separation

Follow-up analyses to the LPA using the comprehensive assessment suggested that the greatest group mean differences found occurred mostly on RAN and the CELF-4 Spanish subtests (Formulating Sentences, Concepts and Following Directions, Word Structure and Recalling Sentences). Moderate group mean differences were estimated on SPELT-3, SNWR and WNV, and small group means were estimated on D and GETU.

The finding of strong or moderate relationships between groups, and the CELF-4 Spanish subtests and SNWR is consistent with previous studies suggesting these measures are sensitive to low language abilities (e.g., Dawson et al., 2003; Kapantzoglou, Fergadiotis, & Restrepo, 2010; Klee et al., 2004; Klee et al., 2007; Owen & Leonard, 2002; Wiig et al., 2006; Windsor et al., 2010). Results are not consistent with studies suggesting grammaticality in language samples as one of the best indicators of children's language abilities (e.g., Restrepo, 1998; Simon-Cereijido & Gutierrez-Clellen, 2007). Perhaps, the combination of measures used for assessment is related to which measures drive the classification. For example, when language sample analysis was used alone as method of assessment in Spanish, grammatical errors per TU was the only indicator of a low group (high scores indicate low performance), which is more consistent with a grammar as a marker of PLI (e.g., Restrepo, 1998; Simon-Cereijido & Gutierrez-Clellen, 2007; Van der Lely, 1998; 2005).

Different solutions and low correlations across measures of the same construct in different tasks suggest that the tasks may be tapping onto different language abilities, that some of the measures are better than others at measuring the constructs, or that the measures tap different aspects of the same construct. Indeed, the correlations between D and the CELF-4 Spanish subtests although significant, they were low, at the .30 level. Low correlations, at the .30 level, were also found between grammatical errors per TU and the CELF-4 Spanish grammatical subtests. Results indicate that lexical diversity in a language sample measures semantic abilities at a different level than the Concept and Following Direction subtest of CELF-4 Spanish given significant but low correlations (.28). Similarly, grammaticality in a language sample taps onto grammatical abilities at a different level than the Recalling Sentences, Word Structure and Formulating Sentences subtests of CELF-4 Spanish.

English language abilities as measured by SPELT-3 were another indicator that showed a strong relationship with group membership, as is indicated when assessing in bilingual populations to identify PLI (e.g., Bedore & Peña, 2008; Bedore et al., 2002; Restrepo & Kruth, 2000). When the comprehensive assessment or CELF-4 Spanish alone were used for assessing language abilities in Spanish, results suggested that English language abilities as measured by SPELT-3 were at the same level as Spanish across groups. In contrast, when Spanish language sample analyses were used, children with average language abilities in Spanish were differentiated into two groups based on SPELT-3: children with relatively low and high levels of English. This solution reflects better the different levels of second language skills that may be observed in bilingual children. Low language performance across both languages could indicate higher probability of diagnosing PLI; nevertheless, in the present solution, given that the average performance of the relatively low-English group is near the mean, different levels in English probably reflect different language proficiency levels in children with typical language development. This is important information for differential diagnosis and indicates that maybe future studies should provide more comprehensive assessment of English skills as well, rather than using a single measure. Regardless, the "low-grammar group" and the "relatively-low and slow group", which may include children with PLI, still demonstrated low English skills consistently with the expected profile of children with PLI.

Non-verbal cognitive skills followed the same pattern as English skills in that they had a positive relationship with language abilities (e.g., Conti-Ramsden et al., 1997; Bishop et al., 1995). When the comprehensive assessment or CELF-4 Spanish was used for assessing language abilities in Spanish, results suggested that non-verbal abilities as measured by WNV were at the same level as Spanish. Thus, there were no groups with same language abilities in Spanish that differed on non-verbal abilities. On the other hand, when language sample analyses were used for Spanish, WNV differentiated children with average Spanish language abilities into two groups based on relatively high and low levels of non-verbal cognitive abilities. The correlational analysis supports the findings of disassociation between the language sample measures and WNV. Perhaps, given the more linguistic nature of measures based on language sample analysis than of those of CELF-4 Spanish, there was greater disassociation of the language and non-verbal cognitive measures, which in turn allowed for more patterns to occur.

Theoretical Implications

The current study by examining latent language ability groups based on semantic and grammatical measures contributes to the literature investigating the relationship between these domains. The nature of the relationship between grammar and lexicon has been discussed for many years as a part of the broader philosophical question regarding what is language and how it develops (e.g., Chomsky, 1970; Locke, 1983, 1997; Pinker & Ullman, 2002; Bates & MacWhinney, 1989). Empirical evidence for the relationship between the two language domains have been obtained based on early language development (e.g., Fenson, Dale, Reznick, Bates, Thal, Pethick, Tomasello, Mervis, & Stiles, 1994; Bates et al., 1994) and children with PLI (e.g., Van der Lely, 1998, 2005; Rice, Wexler, Marquis, & Hershberger, 2000) in English monolingual children and in bilinguals (e.g., Castilla et al., 2009; Convoy & Thal, 2006; Marchman, Martínez-Sussman, & Dale, 2004; Simon-Cereijido & Gutierrez-Clellen, 2009). Most studies suggest strong association between the two domains, although there is also research supporting their relative disassociation based on disproportional semantic and grammatical deficits in children with PLI for example (e.g., Van der Lely, 1998). On the contrary, the current study with LPA, examined directly the patterns of semantic and grammatical abilities in the sample. Results differed within the same sample depending on the measures used: grammatical error per TU and lexical diversity measured with D in language samples suggested disassociation between the two domains given a group with selective grammatical deficits; comprehensive assessment and CELF-4 Spanish did not suggest any disassociation between the two domains. Given that results depended on the type of measures used, more complex models that include these two language domains as latent factors may be more appropriate for addressing related theoretical questions.

Clinical Implications

Results of the present study indicated that there are more than two language ability groups in the general population, and thus dividing children into low and high, may not be consistent with the groups in the population. In this case, diagnosing children with a dichotomy, with and without PLI, could lead to misclassification, depending on the measures used. Groups with performance as low as 2 *SD* from the mean differed mostly on grammatical abilities and processing speed. Importantly, results suggested that not any grammatical measure could capture the low grammatical abilities of some children, only grammatical errors per TU based on language samples revealed a "low-grammar group." Interpretation of the results should be made with caution though, given that the combination of measures influenced the results. Further, English language and non-verbal cognitive measures differentiated two groups of children who scored near average in Spanish, with relatively low and high second language skills, which is consistent with previous studies indicating the need to assess both languages in bilinguals.

Limitations and Future Directions

The current study examined the structure of language ability groups in predominately Spanish-speaking children. Findings were not consistent with the current clinical process of assuming two ability groups in the population, with high and low language abilities. Results suggest more than two distinct language ability groups in bilingual children living in the U.S. Stability in participant classification although it was high considering the level of abilities, results suggested two different language profiles consistent with PLI estimated with two different methods of assessment respectively: the "low-grammar group" based on language sample analysis, and the "relatively-low and slow group" based on comprehensive assessment which included a rapid automatic naming task. Also, children in these two groups were different. Investigation of language ability groups with a combination of language samples analysis and cognitive/processing measures could provide evidence on the stability of the "low-grammar" and the "relatively-low and slow" profiles.

A variety of indicators of language ability were chosen, and they were selected because they have been found to be sensitive to the language-ability differences in bilingual children with PLI. Two tasks, RAN and grammatical errors per TU, identified two low groups with different profiles respectively, and with proportions similar to the prevalence of PLI. Data on children's ability to learn could validate the current finding based on measures at one point in time or could refine group characteristics. For example, measuring children's ability to learn could assist in differentiating children with low abilities in both languages due to limited language experience from those with PLI. Such measures may be less influenced by previous language experience than static measures, and therefore, useful for complementing the assessment of children from culturally and linguistically diverse backgrounds (e.g., Vygotsky, 1935).

A comprehensive assessment in English, similar to the assessment used in Spanish would provide a more accurate description of children's English language abilities. Further, language modality, receptive or expressive, in these different domains may yield groups with more refined characteristics and more stable across methods of assessment. However, in the current study, the standardized measure examined semantic abilities only in the receptive modality and lexical diversity in the language sample analysis was in the expressive modality. In contrast, the grammar construct was only examined in the expressive modality (c.f., Conti-Ramsden et al., 1997).

To minimize dependency on measures and address the relationship of the semantic, grammatical and cognitive domains at the theoretical level, future studies could use factor mixture modeling. Three or more measures for each language domain assessed would allow estimating factors, based on the common variance among the measures, which could be used as group indicators. If so, groups would be less affected by task-specific characteristics, such as in some of the methods in the current study. Nevertheless, factor analysis should precede the estimation of more complex models to decide how measures should be grouped.

Finally, it is possible that groups may vary with age, although Tomblin and Zhang (2006) found that the relationship between underlying semantic and grammatical abilities remain practically the same or underwent very small changes from kindergarten to eighth grade. The current study included children 5-7 years old as a group. Future studies could investigate language ability groups longitudinally to examine age effects on language patterns.

This study is a first step toward the identification of language profiles in bilinguals for more accurate diagnosis. Replication of the current findings with a different population and with older children would strengthen the results on the groups. In bilinguals, the current study may be replicated with children attending bilingual programs, instead of English only education, or have parents from different educational and economical levels.

References

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher*, *33*(1), 4–14.

Algozzine, B., Ysseldyke, J. E., & Christenson, S. (1983). An analysis of the incidence of special class placement: The masses are burgeoning. *The Journal of Special Education*, *17*, 141-147.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed). Washington, DC: American Psychiatric Association.

American Speech-Language-Hearing Association. (1997). Guidelines for Audiologic Screening [Guidelines]. Available from www.asha.org/policy.

Anderson, R. (1996). Assessing the grammar of Spanish-speaking children: A comparison of two procedures. *Language, Speech, and Hearing Services in Schools, 27*, 333–344.

Anderson, R. T. (2004). First language loss in Spanish-speaking children: Patterns of loss and implications for clinical practice. In B. Goldstein (Ed.), *Bilingual language development and disorders in Spanish-English speakers* (pp. 187-211). Baltimore: Brookes.

Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Communication Disorders*, *41*(6), 675-693.

Archibald, L., & Joanisse, M. F. (2009). On the sensitivity and specificity of non-word repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research, 52*, 899–914.

Artiles, A. J., & Ortiz, A. (Eds.). (2002). *English language learners with special needs: Identification, placement, and instruction.* Washington, DC: Center for Applied Linguistics.

Artiles, A. J., Rueda, R., Salazar, J., & Higareda, I. (2005). Within-group diversity in minority disproportionate representation: English language learners in urban school districts. *Exceptional Children, 7*(1), 283–300.

Artiles, A., J., & Kozleski, E., B. (2010). What counts as response and intervention in RTI? A sociocultural analysis. *Psicothema, 22*(4), 949-954.

Bartholomew, D. J. (2007). The three faces of factor analysis. In R. Cudeck and R. MacCullum (Eds.), *Factor Analysis at 100: Historical Developments and Future Directions*. Mahwah: Erlbaum.

Bartolini, U., Arfe, B., Caselli, C. M., Degasperi, L., Deevy, P., & Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: the contribution of clitics and non-word repetition. *International Journal of Language and Communication Disorders, 41*(6), 695-712.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds), *The crosslinguistic study of sentence processing*, pp. 3–76. New York: Cambridge University Press.

Bauer, D. J., & Curran, P., J. (2004). The Integration of Continuous and Discrete Latent Variable Models: Potential Problems and Promising Opportunities. *Psychological Methods, 9*(1), 3–29.

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11*, 1–29.

Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Speech, Language, Hearing Services in Schools*, *36*, 188 – 200.

Bishop, D. V. M. (1982). *Test for Reception of Grammar.* Manchester, England: Author.

Bishop, D. V. M. (1994). Is Specific Language Impairment a Valid Diagnostic Category? Genetic and Psycholinguistic Evidence. *Philosophical Transactions: Biological Sciences: The Acquisition and Dissolution of Language*, *346*(1315), No. 1315, (Oct. 29, 1994), 105-111.

Bishop, D. V. M., North, T., & Donlan, C. (1995). Genetic basis of specific language impairment: evidence from a twin study. *Developmental Medicine and Child Neurology, 37*(1), 56-71.

Brown, T. A. (2006). Confirmatory Factor Analysis for Applied Research. New York, NY: Guilford Press.

Burlingame I, Sussman H, Gillam RB, & Hay J. (2005). An investigation of speech perception in children with SLI on a continuum of formant transition duration. *Journal of Speech, Language, and Hearing Research, 48,* 805–816.

Burton, V. J., & Watkins, R. V. (2007). Measuring word learning: Dynamic versus static assessment of kindergarten vocabulary. *Journal of Communication Disorders, 40*(5), 335-356.

Caesar, L. G., & Kohler, P. D. (2007). The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools, 38,* 190–200.

Campbell, J. M., Bell, S. K., & Keith, L. K. (2001). Concurrent validity of the Peabody Picture Vocabulary Test 3rd Edition as an intelligence and achievement screener for low SES African American children. *Assessment*. *8,* 85-94.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech & Hearing Research, 40*(3), 519-525.

Catts, H.W., Adolf, S.M., Hogan, T.P., & Ellis Weismer, S. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48*, 1378-1396.

Chomsky, C. (1970). Reading, writing, and phonology. Harvard Educational Review, 40, 287-309.

Collins, L. M., & Lanza, S. T., (2010). *Latent class and latent transition analysis for the social, behavioral, and health sciences*. New York, NY: Wiley.

Collins, L. M., & Wugalter, S. E., (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27*(1), 131-157. doi: 10.1207/s15327906mbr2701_8

Conti-Ramsden, G., Crutchley, A., & Botting, N. (1997). The extent to which psychometric tests differentiate subgroups of children with specific language impairment. *Journal of Speech, Language and Hearing Research, 40*, 765-777.

Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development, 77*, 712–735.

Cummins, J. (1988). “Teachers are not miracle workers”: Lloyd Dunn’s call for Hispanic activism. *Hispanic Journal of Behavioral Science, 10*(3), 263–272.

Cullata, B., Page, J. L., & Ellis, J. (1983). Story retelling as a communicative performance screening tool. *Language, Speech, and Hearing Services in Schools. 14*, 66-74.

Castilla, A., Restrepo, M.A. & Perez-Leroux, A. (2009). Individual differences and language interdependence: A study of sequential bilingual development in Spanish-English preschool children. *International Journal of Bilingual Education and Bilingualism*, 1-16.

Dawson, J. I., Stout, C. E., & Eyer, J. A. (2003). *Structured Photographic Expressive Language Test: Third Edition*. DeKalb, IL: Janelle Publications.

Dollaghan, C. A. (2004). Taxometric Analyses of Specific Language Impairment in 3- and 4-Year-Old Children. *Journal of Speech, Language, and Hearing Research, 47,* 464–475.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136–1146.

Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech and Hearing Research, 39,* 643–654.

Duranti, A. (1997) .*Linguistic Anthropology.* Cambridge: Cambridge University Press.

Edwards, J., & Lahey, M. (1996). Auditory lexical decisions of children with specific language impairment. *Journal of Speech and Hearing Research, 39*, 1263–1273.

Elliot, C. D. (1983). *British Ability Scales.* Windsor, England: NFER-Nelson.

Fenson, L., Dale, P., Reznick, J., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59* (Serial No. 242).

Gagné, P. (2006). Mean and Covariance Structure Mixture Models. In G.R. Hancock & R.D. Mueller (Eds.), *Structural Equation Modeling: A second course.* (pp. 197-224). Greenwich, CT: Information Age Publishing.

Gathercole, S., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*, 336–360.

Geiser, C., Lehmann, W., Corth, M., & Eid, M. (2008). Quantitative and qualitative change in children's mental rotation performance. *Learning and Individual Differences, 18,* 419–429.

Gibson, W. A. (1959).Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis, *Psychometrika, 24*, 229-252.

Girbau D, & Schwartz R. (2008). Phonological working memory in Spanish-English bilingual children with and without Specific Language Impairment. *Journal of Communication Disorders. 41*, 124–145.

Goldman, R., & Fristoe, M. (1986). Goldman-Fristoe Test of Articulation. MN: American Guidance Association.

Goldstein, B. A. (2004). *Bilingual language development and disorders in Spanish-English speakers*. Baltimore, Maryland: Paul H. Brookes Publishing Co.

Graf Estes, K., Evans, J. L., & Else-Quest, N.M. (2007). Differences in nonword repetition performance of children with and without Specific Language Impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50*, 177-195.

Gray, S. (2003). Word-learning by preschoolers with specific language impairment: What predicts success? *Journal of Speech, Language & Hearing Research, 46*(1), 56-67.

Gutiérrez-Clellen, V. F. (1996). Language diversity: Implications for assessment. In K. N. Cole, P. S. Dale & D. J. Thal (Eds.). *Assessment of communication and language, Vol. 6*. (pp. 29-56). Baltimore, MD, US: Paul H Brookes Publishing.

Gutiérrez-Clellen, V. F., & Hofstetter, R. (1994). Syntactic complexity in Spanish narratives: A developmental study. *Journal of Speech and Hearing Research, 37*, 645–654.

Gutiérrez-Clellen, V. F., Restrepo, M.A., Bedore, L., Peña, E., & Anderson, R. (2000). Language sample analysis in Spanish-speaking children: Methodological considerations. *Language, Speech, and Hearing Services in Schools, 31*, 88-98.

Gutiérrez-Clellen, V. F., Restrepo, M. A., & Simon-Cereijido, G. (2006). Evaluating the discriminant accuracy of a grammatical measure with Spanish-speaking children. *Journal of Speech, Language, and Hearing Research, 49*, 1209–1223.

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2007). The discriminant accuracy of a grammatical measure with Latino English-speaking children. *Journal of Speech-Language-Hearing Research, 50,* 968-981.

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2009). Using language sampling in clinical assessments with bilingual children: Challenges and future directions. *Seminars in Speech and Language, 30*, 234-245.

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2010). Using Nonword Repetition Tasks for the Identification of Language Impairment in Spanish-English-Speaking Children: Does the Language of Assessment Matter? *Learning Disabilities Research & Practice, 25*(1), 48-58.

Hammer, C. S., Detweiler, J. S., Detweiler, J., Blood, G. W., & Qualls, C. D. (2004). Speech language pathologists' training and confidence in serving Spanish-English bilingual children. *Journal of Communication Disorders, 37,* 91-108.

Hanson R, & Montgomery H (2002). Effects of general processing capacity and sustained selective attention on temporal processing performance of children with specific language impairment. *Applied Psycholinguistics, 23,* 75–93.

Harry, B., & Klingner, J. K. (2006). *Why are so many minority students in special education? Understanding race and disability in schools*. New York : Teachers College Press.

Hicks, B. M., Markon, K. E., Patrick, C. J., Krueger, R. F., & Newman, J. P. (2004). Identifying Psychopathy Subtypes on the Basis of Personality Structure, *Psychological Assessment, 16*(3), 276–288.

Horton-Ikard, R., & Ellis Weismer, S. (2007). A preliminary examination of vocabulary and word learning in African American toddlers from middle and low socioeconomic status homes. *American Journal of Speech-Language Pathology, 16*(4), 381-392.

Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing, 19*(1), 57-84.

Kail, R. (1994). A method for studying the generalized slowing hypothesis in children with specific language impairment. *Journal of Speech and Hearing Research, 37*, 418-421.

Kail, R., & Leonard, L. B. (1986). *Word-finding abilities in language-impaired children.* ASHA Monographs, Number 25.

Kapantzoglou, M., Fergadiotis, G., & Restrepo, M. A. (2010). *Lexical Diversity and Language Sample Elicitation Effects in Spanish-Speaking Children with and without Language Impairment.* Platform presentation presented at the annual conference of the International Association of Logopedics and Phoniatrics, in Athens, Greece.

Kapantzoglou, M., Restrepo, M. A., & Thompson, M.A. (2012). Dynamic assessment of word learning skills: identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools, 43*, 81-96.

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders*, *12*(2), 28-41.

Klee, T., Gavin, W. J., & Stokes, S. F. (2006). Utterance length and lexical diversity in American- and British-English speaking children: What is the evidence for a clinical marker of SLI? In R. Paul (Ed), Language disorders from a developmental perspective: Essays in honour of Robin S. Chapman (pp. 103–140). Mahwah, NJ: Lawrence Erlbaum.

Klee, T., Stokes, S. F., Wong, A. M., Fletcher, P., & Gavin, W. J. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language & Hearing Research, 47*(6), 1396-1410.

Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling* (3rd ed.). New York, NY: Guilford Press.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders, 43*, 465–473.

Kohnert, K. J., & Bates, E. (2002). Balancing bilinguals II: Lexical comprehension and cognitive processing in children learning Spanish and English. *Journal of Speech, Language & Hearing Research, 45*(2), 347-359.

Kohnert, K., Kan, P.F., & Conboy, B.T. (2010). Lexical and grammatical associations in sequential bilingual preschoolers. *Journal of Speech*, Language, and Hearing Research. *53*, 684-698. doi:10.1044/1092-4388(2009/08-0126)

Kohnert, K., Windsor, J., & Pham G. (2009). *Separating differences from disorders using processing-dependent measures*. Poster presented at the Symposium for research in child language disorders, Madison, WI (2009)

Laing, S. P., & Kamhi, A. (2003). Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools, 34*, 44–55.

Lazarsfeld, P.F., & Henry, N.W. (1968). *Latent Structure Analysis*. New New York, NY: Houghton, Mifflin.

Lei, L., Pan, J., Liu, H., McBride-Chang, C., Li, H., Zhang, Y., Chen, L., Tardif, T., Liang, W., Zhang, Z., & Shu, H. (2010). Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. *Journal of Child Psychology and Psychiatry, 52*, 212-220.

Leonard LB. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Leonard, L. (1987). Is specific language impairment a useful construct? In S. Rosenberg (Ed.), *Advances in applied psycholinguistics, Volume 1: Disorders of first-language development*. New York: Cambridge University Press.

Leonard, L. (1991). Specific language impairment as a clinical category. *Language, Speech, and Hearing Services in Schools, 22*, 66-68.

Leonard, L.B., Eyer, J. M., Bedore, L. M., & Grela, B. G. (1997). Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. *Journal of Speech and Hearing Research, 40*, 741-753.

Limbos, M. M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. Journal of Learning Disabilities, 34(2), 136-151.

Lo, Y., Mendell, N.R., & Rubin, D.B. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767-778.

Locke, J.L. (1983). *Phonological acquisition and change*. New York: Academic Press.

Locke, J.L. (1997). A theory of neurolinguistic development. *Brain and Language*, *58*, 265–326.

Lubke, G. H. & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods, 10,* 21-39.

Lubke, G. H. & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling*, *14*(1), 26–47.

Magidson, J. & Vermunt, J.K. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research, 20*, 36-43.

Malvern, *D. D*. & Richards, B.J. 1997: A new measure of lexical diversity. In Ryan, A. and Wray, A., (Eds.), *Evolving models of language (pp. 58-71)*. Clevedon: Multilingual Matters.

Marchman, V., Martínez-Sussman, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. Developmental Science, 7, 212–224.

Marsh, H. W., Hau, K. T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics. A festschrift for Roderick P. McDonald* (pp. 275–340). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralising Hu & Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320–341.

Marsh, W. H., Ludtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical Latent Profile Analysis of Academic Self-Concept Dimensions: Synergy of Person- and Variable-Centered Approaches to Theoretical Models of Self-Concept. *Structural Equation Modeling, 16*, 191–225.

Marshall, C. M., & van der Lely, H. K. J. (2006). A challenge to current models of past tense inflection: the impact of phonotactics. *Cognition, 100*(2), 302–320.

Marton, K., & Schwartz R. G. (2003). Working Memory Capacity Limitations and Language Processes in Children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 46*, 1138-53.

Mayer, M. (1967a). *A boy, a dog, and a frog*. New York: Pied Piper Books.

Mayer, M. (1967b). *Frog on his own*. New York: Pied Piper Books.

McGregor, K. K., Friedman, R. M., Reilly, R. M., & Newman, R. M. (2002). Semantic representations and naming in young children. *Journal of Speech, Language, and Hearing Research, 45*, 332–346.

McKee, G., Malvern, *D*., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing, 15*(3), 323-337.

McKenna, P., & Warrington, E.K. (1983). *The Graded Naming Test.* Windsor, UK: NFER-Nelson.

McLachlan, G. J., & Basford, K. E. (1988). Mixture Models. New York, NY: Marcel Dekker.

Miller, C., Kail, R., Leonard, L., & Tomblin, J. B. (2001). Speed of processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*, 416-433.

Miller, J. F., & Iglesias, A. (2008). Systematic analysis of language transcripts (SALT), Bilingual SE Version 2008 [Computer software]. SALT Software, LLC.

Montgomery, J. (2000). Verbal working memory in sentence comprehension in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 43*, 293–308.

Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition,* 5(1), 39-68.

Morgan, G., Restrepo, M. A., Auza, A. (in press). Incomplete Acquisition of Spanish Morphology: Impact on the identification of Language Impairment in Spanish-English bilingual children. *Journal of Speech, Language, and Hearing Research.*

Morgan, G., Srivastava, P., Restrepo, M.A.,& Auza, A (2009). *Rapid Automatic Naming in Bilingual Spanish-English and Monolingual Spanish-speaking children with and without Language Impairments*. Poster presentation at the Symposium on Research in Child Language Disorders.

Muthén B, Muthen L. (1998-2010). Mplus User's Guide. Los Angeles, CA: Muthen & Muthen.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New Developments and Techniques in Structural Equation Modeling* (pp. 1-33). Lawrence Erlbaum Associates.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 345-368). Newbury Park, CA: Sage Publications.

Muthén, B. & Muthen, L. (2000). Integrating person-centered and variable-centered analysis: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24,* 882-891.

Myung, J. I. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology, 47*, 90-100.

National Research Council (2002). Scientific re-search in education. R. Shavelson & L. Towne (Eds.), Committee on Scientific Principles for Educational Research. Washington, DC: National Academy Press.

Newcomer, P. L., & Hammill, D. D. (1988). *Test of Language Development-2 Primary.* Austin, TX: Pro-Ed.

Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology, 90*(2), 142–171.

Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*, 207–223.

Owen, A. J., & Leonard, L. B. (2002). Lexical diversity in the spontaneous speech of children with specific language impairment: Application of *D. Journal of Speech, Language & Hearing Research, 45*(5), 927.

Owen, A. J., & Leonard, L. B. (2006). The production of finite and nonfinite complement clauses by children with specific language impairment and their typically developing peers. *Journal of Speech, Language, and Hearing Research, 49*, 548–571.

Palladino, P., & Cornoldi, C. (2004). Working memory performance in Italian students with foreign language learning difficulties. *Learning and Individual Differences, 14*, 137-151.

Paradis, J. & Crago, M. (2000). Tense and temporality: Similarities and differences between language-impaired and second-language children. *Journal of Speech, Language, and Hearing Research, 43*, 834–848.

Paradis, J. & Crago, M. (2004). Comparing L2 and SLI grammars in French: Focus on DP. In Prévost, P. & Paradis, J. (Eds.). *The acquisition of French in different contexts: Focus on functional categories* (pp. 89–108) Amsterdam: John Benjamins.

Paradis, J., Rice, M., Crago, M., & Marquis, J. (2008). The acquisition of tense in English: Distinguishing child L2 from L1 and SLI. *Applied Psycholinguistics, 29*, 1–34.

Pastor, D. A., Barron, K. E., Miller, B. J., & Davis, S. L. (2007). A latent profile analysis of college students' achievement goal orientation profiles. *Contemporary Educational Psychology, 32*, 8-47.

Pearson, B. Z., Fernández, S. C., Lewedeg, V., & Oller, K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics, 18*(1), 41-58.

Peña, E. D., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word-learning ability. *American Journal of Speech-Language Pathology, 10*(2), 138-154.

Peterson, R. A. (2000). *Constructing effective questionnaires.* Thousand Oaks, CA: Sage.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, *6*(11), 456–463.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*, 879-903.

Raykov, T. & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. New York, NY: Taylor and Francis.

Rapin, I., & Allen, D. A. (1987). Developmental dysphasia and autism in preschool children: Characteristics and subtypes. In J. Martin, P. Fletcher, R. Grunwell, & D. Hall (Eds.), *Proceedings of the first international symposium on specific speech and language disorders in children* (pp. 20–35). London: Afasic.

Raven, J. C. (1986). *Coloured progressive matrices*. London: H. K. Lewis & Co.

Redmond, S.M. (2003). Children's productions of the affixed in past tense and past participle contexts. *Journal of Speech, Language, and Hearing Research, 46*, 1095-1109.

Renfrew, C. (1991). *The Bus Story: A test of continuous speech*. Oxford, England: C. E. Renfrew.

Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. Journal of Speech and Hearing Disorders, 54, 587-599.

Rescorla, L, & Achenbach, T. M. (2002). Use of the language development survey in a national probability sample of children 18 to 35 months old. *Journal of Speech, Language, and Hearing Resrarch, 45*, 733-743.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research, 41*(6), 1398-1411.

Restrepo, M.A., Castilla, A.P., Arboleda, A., Schwanenflugel, P., Neuhart Prittchett, S., & Hamilton, C. (2010). Sentence length, complexity and grammaticality growth in Spanish-speaking children attending English-only and bilingual preschool programs. *Language Speech and Hearing Services in Schools, 41,* 3-13.

Restrepo, M. A., & Kruth, K. (2000). Grammatical characteristics of a bilingual student with specific language impairment. *Journal of Children's Communication Development, 21*, 66–76.

Restrepo, M. A. & Silverman, S. (2001). The validity of the Spanish Preschool Language Scale for use with bilingual children. *American Journal of Speech-Language Pathology, 10,* 382-393.

Restrepo, M.A., Schwanenflugel, P., Blake, J., Neuhart-Pritchett, S., Cramer, S., & Ruston (2006). Performance on the PPVT-III and the EVT: Applicability of the Measures with African American and European American preschool children. *Language, Speech, and Hearing Services in Schools*. *37*, 17-27.

Restrepo, M. A., & Gutierrez-Clellen, V. F. (2001). Article Use in Spanish-speaking children with SLI. *Journal of Child Language*, *28*, 433-452.

Restrepo, M. A., Kapantzoglou, M., Gorin, J., & Gray, S. (2011). *Phase II Development of a Language Screener for Bilingual Spanish Speaking Children*. The Fifth MG & WGs Meeting. Cost Action IS0804, Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment. Malta.

Restrepo, M. A., Swisher, L., Plante. E., & Vance, R. (1992). Relations among verbal and

nonverbal cognitive skills in normal language and specifically language impaired children. *Journal of Communication Disorders*. *25*, 205-219.

Reuterskiöld-Wagner, C., Sahlén, B., & Nyman, A. (2005). Non-word repetition and non-word discrimination in Swedish preschool children. *Clinical Linguistics and Phonetics, 34*, 337-352.

Reynolds, C., R. (1982). The problem of bias in psychological assessment. In C.R. Reynolds, T.B. Gutkin (Eds.), *The handbook of school psychology*, (pp. 178-208). New York, NY: Wiley.

Rice, M. L., Wexler, K., Marquis, J., & Hershberger, S. (2000). Acquisition of irregular past tense by children with SLI. *Journal of Speech, Language, and Hearing Research*, *43*, 1126–1145.

Rodekohr, R. K., & Haynes, W. O. (2001). Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders, 34,* 255–272.

Roseberry, C., & Connell, P. (1991). Use of an invented language rule in the differentiation of normal and specific language-impaired Spanish-speaking children. *Journal of Speech and Hearing Research, 34*, 596-603.

Roseberry-McKibbin, C., Brice, A., & O'Hanlon, L. (2005). Serving English language learners in school settings: A national survey. *Language, Speech, and Hearing Services in Schools, 36,* 48–61.

Rost, J. (2006). Latent-class-analyses. In F. Petermann & M. Eid (Eds.), *Handbuch der psychologischen Diagnostik*, (pp. 275–287). Gottingen: Hogrefe.

Samson, J. F., & Lesaux, N. K. (2009). Language-minority learners in special education: Rates and predictors of identification for services. *Journal of Learning Disabilities, 42*(2), 148-162.

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal American Statistical Association*. 81 826-831.

Schiff-Myers, N. (1992). Considering arrested language development and language loss in the assessment of second language learners. *Language, Speech, and Hearing Services in Schools, 23,* 28–33.

Schwartz, R. G. (2009). *Handbook of child language disorders*. New York: Psychology Press.

Seiger-Gardner, L., & Brooks, P. J. (2008). Effects of onset- and rhyme-related distractors on phonological processing in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 51*(5), 1263-1281.

Semel, E., Wiig, E. H., & Secord, W. (1980). *Clinical evaluation of language fundamentals-revised.* San Antonio, TX: Psychological Corporation.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research, 4*(6), 1461–1481.

Simon-Cereijido, G. & Gutiérrez-Clellen, V.F. (2007) Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics, 28,* 317-339.

Simon-Cereijido, G. & Gutiérrez-Clellen, V.F. (2009) A cross-linguistic and bilingual evaluation of the interdependence between lexical and grammatical domains. *Applied Psycholinguistic*. 30: 315-338.

Skiba, R., Simmons, A., Ritter, S., Kohler, K., Henderson, M., & Wu, T. (2006). The Context of Minority Disproportionality: Perspectives on Special Education Referral. *Teachers College Record, 108*(7), 1424-1459.

Smyk, E., Restrepo, M. A., Gorin, J., & Kapantzoglou, M. (2009). *A global rating scale of language proficiency in Spanish-English bilingual children*. Paper session presented at the annual meeting of the American Speech-Language-Hearing Association, New Orleans, LA.

Smyk, E., Restrepo, M.A., Morgan, G., & Kapantzoglou, M. (2009). *Development of parallel measures for story retelling and comprehension tasks for Spanish-speaking first graders*. Poster session presented at the annual Symposium on Research in Child Language Disorders, Madison, WI.

Solórzano, R. W. (2008). High Stakes Testing: Issues, Implications, and Remedies for English Language Learners. *Review of Educational Research*, 2008, 78: 260.

Spearman, C. (1904). General intelligence objectively determined and measured, *American Journal of Psychology, 15*, 201-293.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*(6), 1321-1337.

Tallal P, Piercy M. (1973). Defects of non-verbal auditory perception in children with developmental aphasia. *Nature,* 241, 468–9.

Tallal P, Piercy M. (1974). Developmental aphasia: rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia, 12,* 83–93.

Thal, D., O'Hanlon, L., Clemmons, M., & Frailin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language and Hearing Disorders, 42*, 482-496.

Thomas, M.L., Lanyon, R.I., & Millsap, R.E. (2009). Validation of diagnostic measures based on latent class analysis: A step forward in response bias research. Psychological Assessment, 21, 227-230.

Thordardottir, E. T., & Namazi, M. (2007). Specific language impairment in French-speaking children: Beyond grammatical morphology. *Journal of Speech, Language, and Hearing Research, 50*(3), 698-714.

Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N., & Chilingaryanb, G. (2011). Sensitivity and Specificity of French Language and Processing Measures for the Identification of Primary Language Impairment at Age 5. *Journal of Speech, Language, and Hearing Research, 54,* 580–597.

Tofighi, D., & Enders, C. K. (2007). Identifying the correct number of classes in a growth mixture models. In G. R. Hancock & K. M. Samuelson (Eds.), Advances in latent variable mixture models (pp. 317–341). Greenwich, CT: Information Age.

Tomblin, B. J., Records, N. L., Buckwalter, P., Zhang, X., & Smith, E. (1997). Prevalence of Specific Language Impairment in Kindergarten Children. *Journal of Speech, Language, and Hearing Research*. *40*, 1245-1260.

Tomblin, J. B., & Zhang, X. (1999). Language patterns and etiology in children with specific language impairment. In H. Tager-Flusberg (ed), *Neurodevelopmental disorders,* (pp. 361-482). Cambridge, MA: MIT Press.

Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, *32*(5), 323-352.

van de Vijver, F .J. R., & Tanzer, N. K., (2004). Bias and equivalence in cross-cultural assessment: an overview. *Revue Européenne de Psychologie Appliquée 54* (2004) 119–135.

van der Lely, H. K. J. (1998) SLI in children: Movement, economy and deficits in the computational syntactic system. *Language Acquisition 72,* 161–192

van der Lely, H. K. J. (2005). Domain-specific cognitive systems: Insight from Grammatical specific language impairment. *Trends in Cognitive Sciences, 9,* 53–59.

Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing, 17*(1), 65-83.

Vitevitch, M. S. & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes, 21*(6), 760-770.

Vygotsky, L. S. (1935). The Dynamics of the Schoolchild's Mental Development in Relation to Teaching and Learning. *Journal of Cognitive Education and Psychology, 10*(2), 2011, 198-211.

Watkins, R. V., Kelly, *D*. J., Harbers, H. M., & Hollis, W. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech and Hearing Research, 38*, 1349-1355.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. New York: Psychological Corporation.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence - III*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2005). *Wechsler Intelligence Scale for Children—Fourth Edition Spanish.* San Antonio, TX: Psychological Corporation.

Wechsler, D., & Naglieri, J.A. (2006). *Wechsler Nonverbal Scale of Ability*. San Antonio, TX--Harcourt Assessment.

Weismer, E., Evans, J., & Hesketh, L. (1999). An examination of verbal working memory capacity in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1249–1260.

Wiig, E. H., Secord, W. A., & Semel, E., (2006). *Clinical Evaluation of Language Fundamentals-4, Spanish Edition*. San Antonio, TX: Pearson.

Wilson, B. C., & Risucci, D. A. (1986). A model for clinical-quantitative classification: Generation 1. Application to language disordered preschool children. *Brain and Language*, *27*, 281-310.

Windsor, J., & Hwang, M. (1999). Testing the generalized slowing hypothesis in specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1205-1218.

Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology, 19*, 298–310.

Windsor, J., Kohnert, K., Loxtercamp, A. L., & Kan, P. (2008). Performance on nonlinguistic visual tasks by children with language impairment. *Applied Psycholinguistics*, *29*(2), 237-268.

Yang, C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis*, *50*(4), 1090–1104.

Ysseldyke; J. E., Vanderwood, M. L., & Shriner, J. (1997). Changes over the past decade in special education referral to placement probability: An incredibly reliable practice. *Diagnostique*, *23*, 193-201.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). *Preschool Language Scale -4: Spanish Edition.* San Antonio, TX: Psychological.

Tables

Table 1
Descriptive Statistics of the Major Study Variables before the Removal of Outliers

Lexical Diversity Index	*n*	*M*	*SD*	Range
CConc	429	16.78	8.77	0.00 - 44.00
CWStr	429	14.36	5.85	1.00 - 32.00
CRecS	429	22.73	17.11	0.00 - 80.00
CFormS	429	14.01	8.65	0.00 - 39.00
D	387	20.99	5.84	8.00 - 41.21
GETU	390	0.15	0.15	0.00 - 1.14
SNWR	415	51.83	10.20	0.00 - 64.00
RAN	429	0.02	0.06	0.00 - 0.71
SPELT	425	21.33	10.89	0.00 - 48.00
WNV	431	25.02	5.85	7.00 - 39.00

Note. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; D=Lexical Diversity; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

Table 2

Descriptive Statistics of the Major Study Variables after the Removal of Outliers

Lexical Diversity Index	*n*	*M*	*SD*	Range
CConc	429	16.78	8.77	0.00 - 44.00
CWStr	429	14.36	5.85	1.00 - 32.00
CRecS	429	22.73	17.11	0.00 - 80.00
CFormS	429	14.01	8.65	0.00 - 39.00
D	387	20.99	5.84	8.00 - 41.21
GETU	385	0.14	0.13	0.00 - 0.75
SNWR	410	52.44	8.59	0.00 - 64.00
RAN	422	0.02	0.04	0.00 - 0.25
SPELT	425	21.33	10.89	0.00 - 48.00
WNV	431	25.02	5.85	7.00 - 39.00

Note. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; D=Lexical Diversity; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

Table 3

Patterns of Missing Data Before and After the Removal of Outliers for Each Indicator

Indicator	Before		After	
	Number Missing	Percentage Missing	Number Missing	Percentage Missing
CConc	2	0.46%	2	0.46%
CWStr	2	0.46%	2	0.46%
CRecS	2	0.46%	2	0.46%
CFormS	2	0.46%	2	0.46%
D	41	9.51%	44	10.21%
GETU	44	10.21%	46	10.67%
SNWR	15	3.48%	21	4.87%
RAN	1	0.23%	9	2.10%
SPELT	6	1.39%	6	1.40%
WNV	0	0.00%	0	0.00%
	Totals			
	115	2.67%	134	3.11%

Note. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; D=Lexical Diversity; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

Table 4

Correlation Matrix of the Measures Used as Group Indicators

Variable	1	2	3	4	5	6	7	8	9
1. CConc	1.00								
2. CWStr	0.64**	1.00							
3. CRecS	0.60**	0.70**	1.00						
4. CFormS	0.71**	0.71**	0.66**	1.00					
5. D	0.28**	0.36**	0.34**	0.33**	1.00				
6. GETU	-0.16**	-0.33**	-0.30**	-0.23**	-0.10	1.00			
7. SNWR	0.44**	0.44**	0.46**	0.47**	0.20**	-0.18**	1.00		
8. RAN	-0.16**	-0.18**	-0.10*	-0.15**	-0.08	0.12*	-0.20**	1.00	
9. SPELT	0.50**	0.39**	0.28**	0.52**	0.23**	-0.11**	0.41**	-0.07	1.00
10. WNV	0.51**	0.29**	0.24**	0.42**	0.08	-0.07	0.32**	-0.13**	0.44**

Note. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; D=Lexical Diversity; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

Table 5

Fit Statistics for the Two-, Three-, Four-, and Five-Group Solutions Based on a Comprehensive Assessment

Model	Free Parameters	SSA-BIC	Entropy	Adjusted LMR-LRT
Two Groups	31	10090.61	0.87	
2Δ*log*L				995.58**
Three Groups	42	9780.28	0.91	
2Δ*log*L				337.10*
Four Groups	53	9512.18	0.90	
2Δ*log*L				295.49**
Five Groups	64	9407.59	0.91	
2Δ*log*L				134.40

Note. 2Δ*log*L = 2 Times the Loglikelihood Difference; SSA-BIC = Sample Size Adjusted Bayesian Information Criterion; LMR-LRT = Lo-Mendell-Rubin Likelihood Ratio Test.

*p < .05. **p < .01.

Table 6

Group Characteristics for the Two-, Three-, Four-, and Five-Group Solutions Based on a Comprehensive Assessment

Model	Group Proportion (*N*=431)	Group Description	Average Group Assignment Probabilities
Two Groups			
Group One	42.20% (182)	High	0.95
Group Two	57.80% (249)	Relatively Low	0.97
Three Groups			
Group One	52.30% (225)	Relatively Low	0.96
Group Two	40.40% (175)	High	0.95
Group Three	7.30% (31)	Relatively-Low and Slow	0.97
Four Groups			
Group One	41.60% (180)	Relatively Low	0.96
Group Two	7.10% (30)	Relatively-Low and Slow	0.98
Group Three	37.90% (163)	Average	0.93
Group Four	13.40% (58)	High	0.94
Five Groups			
Group One	41.00% (177)	Relatively Low 1	0.96
Group Two	37.70% (163)	Average	0.93
Group Three	6.60% (29)	Relatively Low 2	0.99
Group Four	1.20% (5)	Relatively Low 3	1.00
Group Five	13.30% (57)	High	0.93

Table 7
Means and Standard Deviations Measures in the Four-Group Solution Based on a Comprehensive Assessment

	Group 1 *N* = 181		Group 2 *N* = 30		Group 3 *N* = 163		Group 4 *N* = 58	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
CConc	10.25	0.44	11.40	1.14	20.28	0.79	30.10	1.28
CWStr	10.25	0.38	10.54	1.04	16.84	0.49	22.24	0.70
CRecS	12.29	0.94	14.97	2.88	26.12	1.32	49.82	4.19
CFormS	7.33	0.44	9.08	1.09	17.30	0.82	28.17	1.11
D	19.11	0.44	20.08	0.98	21.47	0.50	25.27	1.11
GETU	0.17	0.01	0.21	0.04	0.13	0.01	0.09	0.01
SNWR	48.46	0.83	48.41	1.91	55.36	0.52	59.03	0.56
RAN	0.01	0.00	0.13	0.01	0.01	0.00	0.01	0.00
SPELT	15.76	0.79	18.61	1.74	24.39	1.12	31.23	1.54
WNV	22.28	0.45	22.97	0.97	26.86	0.57	29.41	0.60

Note. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; D=Lexical Diversity; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

Table 8

ANOVA Results for Group Effect on the Indicators of the Comprehensive Assessment

Variable	*df*	*F*	η^2
RAN	3	393.50	0.74
CFormS	3	322.09	0.69
CConc	3	263.49	0.65
CWStr	3	174.09	0.55
CRecS	3	154.48	0.52
SPELT	3	50.03	0.26
SNWR	3	43.75	0.24
WNV	3	41.97	0.23
D	3	17.57	0.12
GETU	3	11.02	0.08

Note. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; D=Lexical Diversity; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability. All comparisons were significant at the $p < 0.001$ level.

Table 9

Fit Statistics for the Two-, Three-, and Four-Group Solutions Based on the CELF-4 Spanish

Model	Free Parameters	SSA-BIC	Entropy	Adjusted LMR-LRT
Two Groups	19	6490.35	0.86	
$2\Delta log$L				837.89**
Three Groups	26	6218.13	0.88	
$2\Delta log$L				285.74**
Four Groups	33	6147.81	0.86	
$2\Delta log$L				88.48

Note. $2\Delta log$L = 2 Times the Loglikelihood Difference; *CELF-4* = Clinical Evaluation of Language Fundamentals-Fourth Edition; SSA-BIC = Sample Size Adjusted Bayesian Information Criterion; LMR-LRT = Lo-Mendell-Rubin Likelihood Ratio Test.

**$p < .01$.

Table 10

Group Characteristics for the Two-, Three-, and Four-Group Solutions Based on the CELF-4 Spanish

Model	Group Proportion (*N*=431)	Group Description	Average Group Assignment Probabilities
Two Groups			
Group One	58.70% (253)	Relatively Low	0.97
Group Two	41.30% (178)	High	0.94
Three Groups			
Group One	47.10% (203)	Relatively Low	0.96
Group Two	39.10% (168)	Average	0.93
Group Three	13.80% (60)	High	0.92
Four Groups			
Group One	46.60% (201)	Relatively Low	0.96
Group Two	11.60% (50)	Average 1	0.95
Group Three	11.70% (50)	Average 2	0.85
Group Four	30.10% (130)	High	0.89

Note. *CELF-4* = Clinical Evaluation of Language Fundamentals-Fourth Edition.

Table 11
Means and Standard Deviations for Measures on the Three-Group Solution Based on the CELF-4 Spanish

	Group 1 *N*=204		Group 2 *N*=167		Group 3 *N*=60	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
CConc	10.13	0.39	20.17	0.69	29.90	1.26
CWStr	10.12	0.34	16.73	0.45	22.16	0.63
CRecS	11.96	0.81	26.33	1.26	49.32	3.69
CFormS	7.25	0.40	17.15	0.69	28.19	1.00
SPELT	15.92	0.73	24.24	1.02	31.26	1.45
WNV	22.23	0.41	26.76	0.52	29.54	0.57

Note. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; D=Lexical Diversity; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

Table 12

ANOVA Results for Group Effects based on CELF-4 Spanish

Variable	*df*	*F*	η^2
CConc	2	390.22	0.65
CWStr	2	283.40	0.57
CRecS	2	249.34	0.54
CFormS	2	575.35	0.73
SPELT	2	77.72	0.27
WNV	2	67.51	0.24

Note. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability. All comparisons were significant at the $p < 0.001$ level

Table 13
Fit Statistics for the Two-, Three-, and Four-Group Solutions Based on Language Samples Analyses

Model	Free Parameters	SSA-BIC	Entropy	Adjusted LMR-LRT
Two Groups	13	4386.07	0.91	
2Δ*log*L				117.87**
Three Groups	18	4296.92	0.72	
2Δ*log*L				100.32**
Four Groups	23	4276.67	0.73	
2Δ*log*L				33.60

Note. 2Δ*log*L = 2 Times the Loglikelihood Difference; SSA-BIC = Sample Size Adjusted Bayesian Information Criterion; LMR-LRT = Lo-Mendell-Rubin Likelihood Ratio Test.
**$p < .01$.

Table 14

Group Characteristics for the Two-, Three-, and Four-Group Solutions Based on Language Sample Analyses

Model	Group Proportion (*N*=431)	Group Description	Average Group Assignment Probabilities
Two Groups			
Group One	7.80% (33)	Low-Grammar	0.97
Group Two	92.20% (398)	Average	0.94
Three Groups			
Group One	6.20% (27)	Low-Grammar	0.96
Group Two	50.00% (215)	Average Spanish & Relatively Low English	0.93
Group Three	43.80% (189)	Average Spanish & Relatively High English	0.92
Four Groups			
Group One	42.30% (182)	Average 1	0.96
Group Two	42.40% (183)	Average 2	0.95
Group Three	2.10% (9)	Relatively Low 1	0.85
Group Four	13.30% (57)	Relatively Low 2	0.89

Table 15

Means and Standard Deviations for Measures on the Three-Group Solution Based on the Language Sample Assessment

	Group 1 *N*=224		Group 2 *N*=176		Group 3 *N*=31	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*
D	31.39	2.52	19.96	0.45	22.64	0.59
GETU	0.01	0.00	0.13	0.01	0.11	0.01
SPELT	58.70	7.21	13.97	1.24	29.90	0.92
WNV	25.44	2.45	22.20	0.51	28.31	0.63

Note. GETU = Grammatical Errors Per T-Unit; SPELT = Structured Photographic Expressive Language Test –Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

Table 16

ANOVA Results for Group Effects on the Indicators for Language Sample Analysis

Variable	*df*	*F*	η^2
D	2	14.97	0.07
GETU	2	189.23	0.50
SPELT	2	321.67	0.60
WNV	2	93.77	0.30

Note. D = Lexical Diversity; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; SPELT = Structured Photographic Expressive Language Test –Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

Table 17

Children's Classification Stability across the Three Methods of Assessment

	CELF – 4 Spanish			Language Sample Analysis		
Group	Group 1 Relatively Low	Group 2 Average	Group 3 High	Group 1 Low-Grammar	Group 2 Average Spanish & Relatively Low English	Group 3 Average Spanish & Relatively High English
Comprehensive Assessment						
Group 1 Relatively Low	97.79%	2.21%	0.00%	6.08%	77.35%	16.57%
Group 2 Relatively-Low and Slow	76.67%	23.33%	0.00%	10.00%	63.33%	26.67%
Group 3 Average	2.44%	94.51%	3.05%	4.27%	34.76%	60.98%
Group 4 High	0.00%	0.64%	99.36%	0.00%	2.56%	97.44%
CELF - 4 Spanish						
Group 1 Relatively Low	x	x	x	5.88%	76.96%	17.16%
Group 2 Average	x	x	x	5.39%	35.33%	59.28%
Group 3 High	x	x	x	0.00%	6.67%	93.33%

Figures

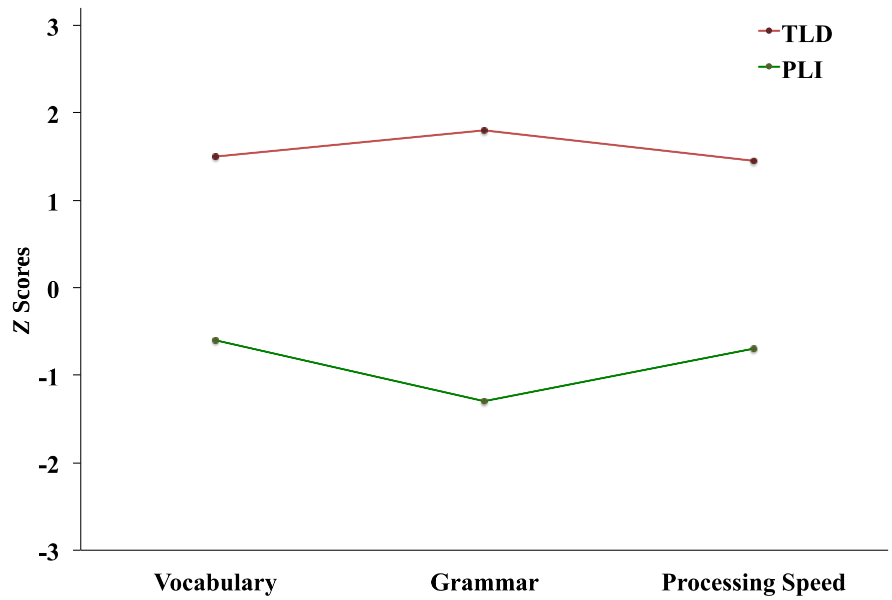


Figure 1. *An example of the average performance of groups across tasks when assuming two groups, with and without language impairment.*

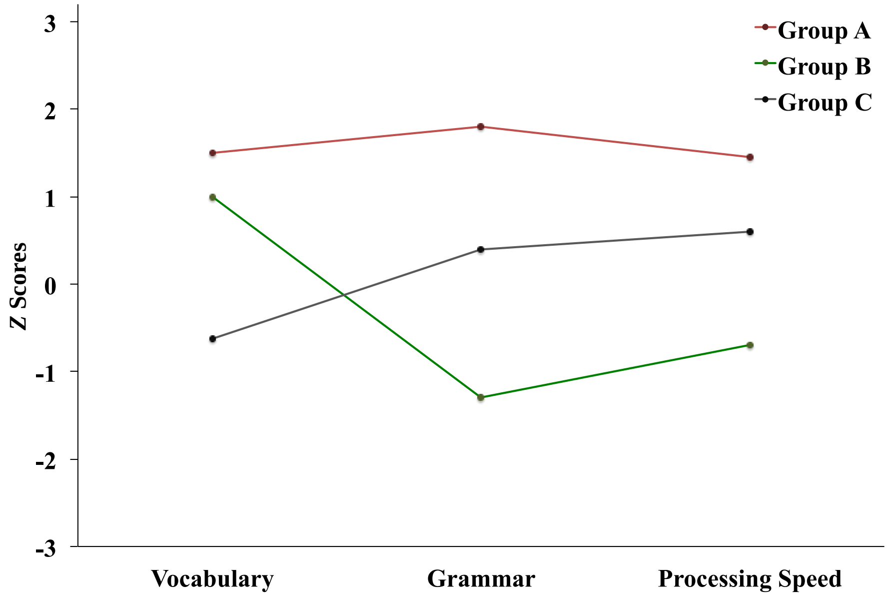


Figure 2. An example of the average performance of three groups with distinct across tasks. Different strengths and weaknesses in this case indicate different types of language area affected for each one of the groups.

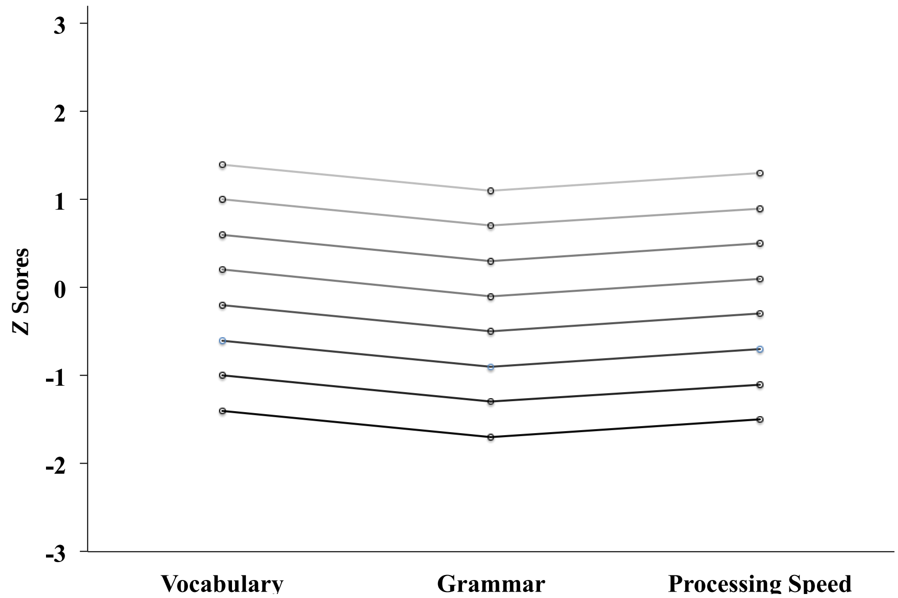


Figure 3. *An example of the average performance across tasks when there are no distinct groups.*

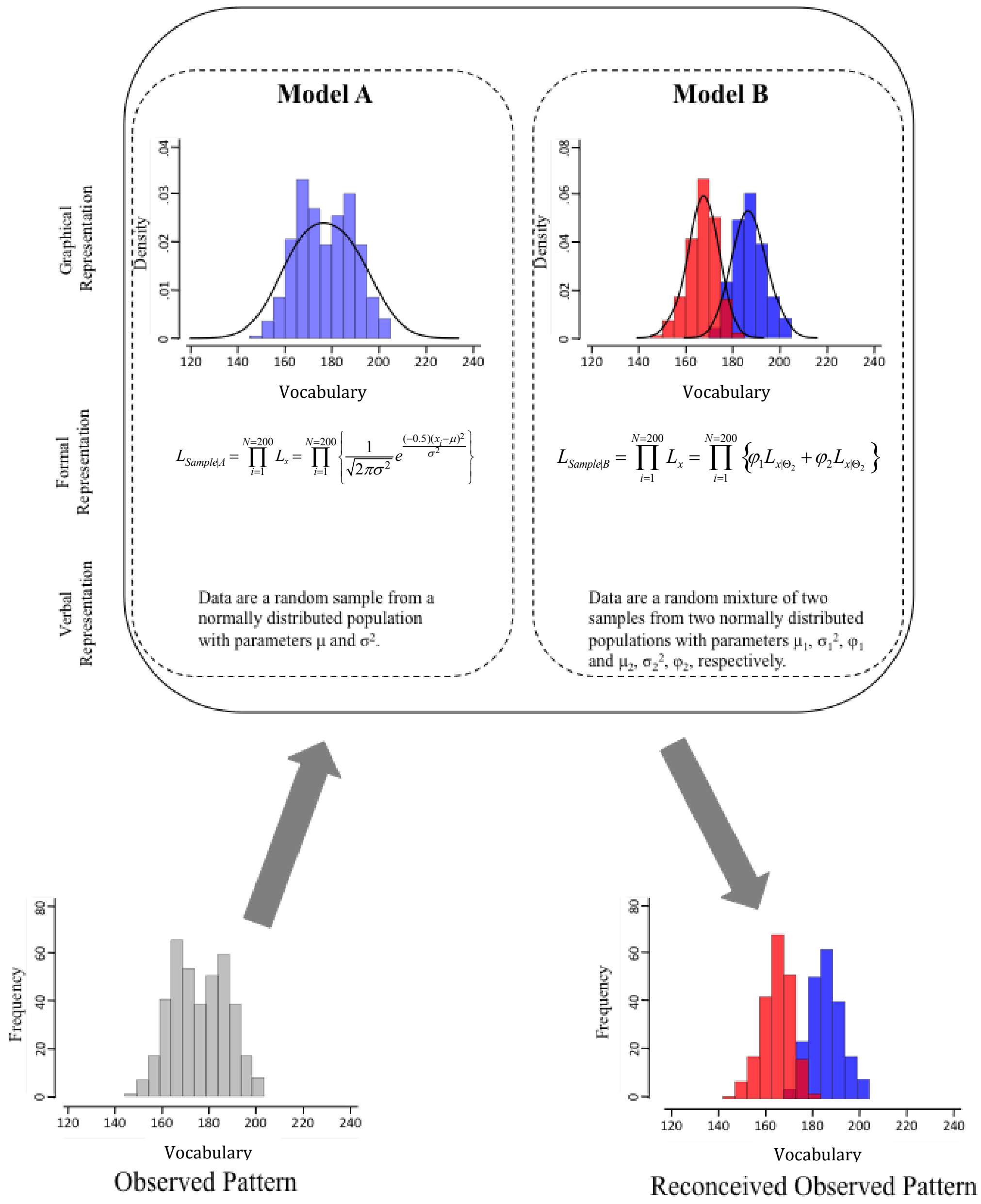
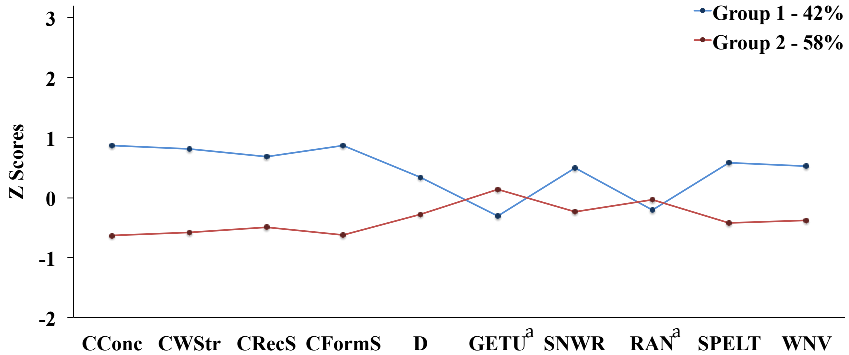
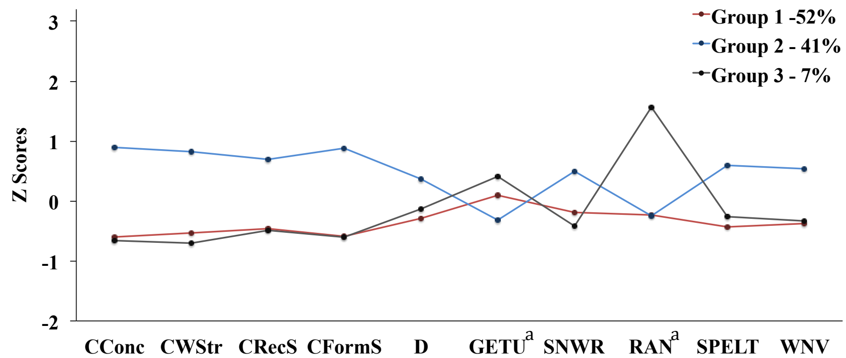


Figure 4. An example of model-based reasoning.

A.



B.



C.

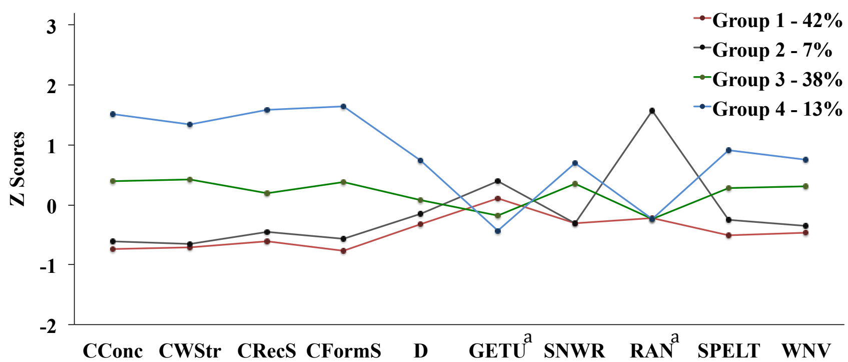


Figure 5. Two-, three-, and four-group solutions for a comprehensive language assessment. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability. [a]High scores reflect low performance.

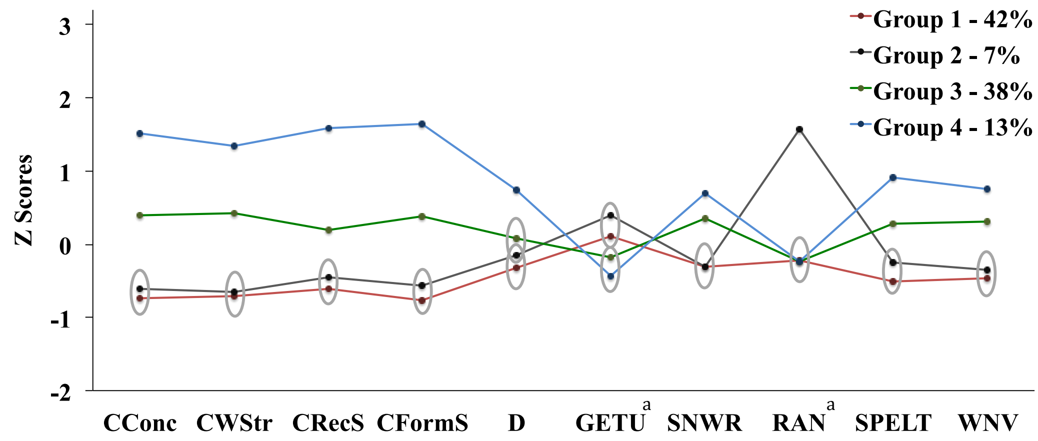
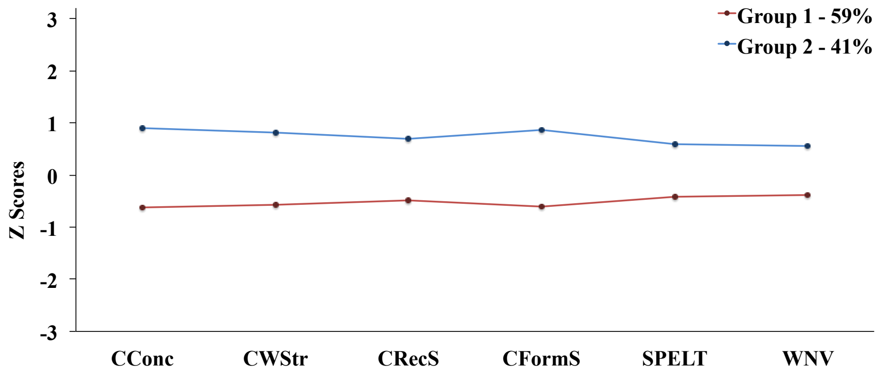


Figure 6. Four-group solution for the comprehensive assessment showing pairwise comparison results. Means in the same ellipses are not significantly different. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; GETU = Grammatical Errors Per T-Unit; SNWR = Spanish Non-word Repetition; RAN = Rapid Automatic Naming; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability. [a]High scores reflect low performance.

A.



B.

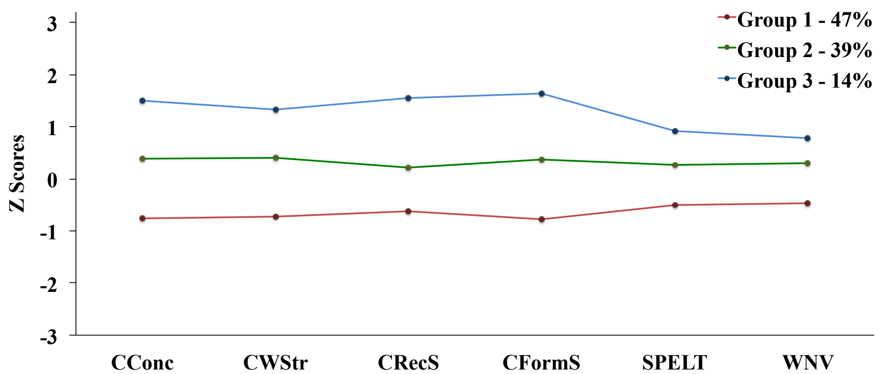
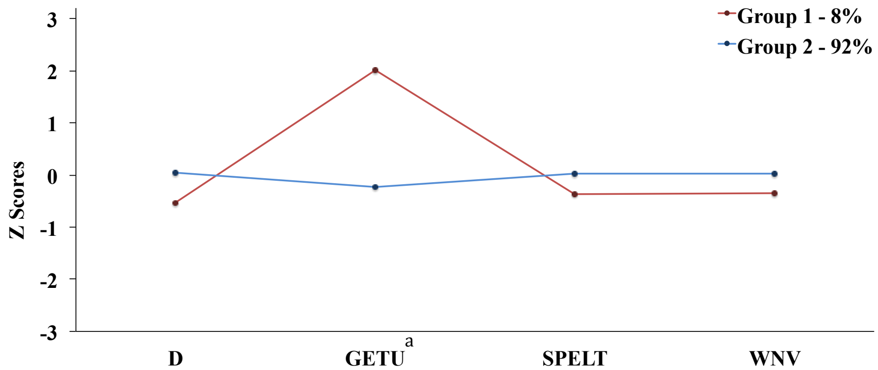


Figure 7. Two- and three-group solutions for the CELF-4 Spanish. CConc = Clinical Evaluation of Language Fundamentals -Fourth Edition Concepts and Following Directions; CWStr = Clinical Evaluation of Language Fundamentals-Fourth Edition Word Structure; CRecs = Clinical Evaluation of Language Fundamentals-Fourth Edition Recalling Sentences; CFormS = Clinical Evaluation of Language Fundamentals-Fourth Edition Formulating Sentences; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability.

A.



B.

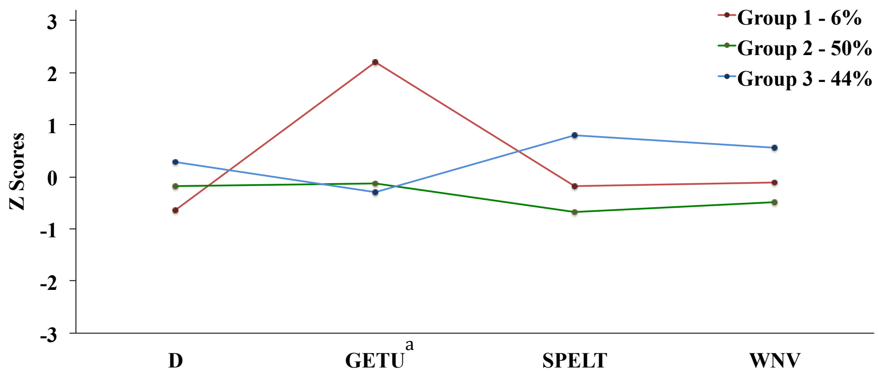


Figure 8. Two- and three-group solutions based on language sample analyses. GETU = Grammatical Errors Per T-Unit; SPELT = Structured Photographic Expressive Language Test -Third Edition; WNV = Wechsler Nonverbal Scale of Ability. [a]High scores reflect low performance.