I, Rochel Lazewnik, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Communication Sciences and Disorders.

It is entitled:
Identifiers of Bilingual Spanish-English Speaking Children with Language Impairment

Student's name: Rochel Lazewnik

This work and its defense approved by:

Committee chair: Nancy Creaghead, Ph.D.
Committee member: Allison Breit-Smith, Ph.D.
Committee member: Lisa Cahill, Ph.D.
Committee member: Jo-Anne Prendeville, Ed.D.
Committee member: Lesley Raisor, Ph.D.
Identifiers of Bilingual Spanish-English Speaking Children with Language Impairment

A dissertation submitted to the Division of Research and Advanced Studies of the University of Cincinnati

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In the Department of Communication Sciences and Disorders in the College of Allied Health Sciences

2013

By

Rochel Lazewnik

B.A., Thomas Edison State College, 2002

M.S., Long Island University, 2005

Committee Chair: Nancy A. Creaghead, Ph.D.

Committee: Lisa Cahill, Ph.D.

Allison Breit Smith, Ph.D.

Jo-Anne Prendeville, Ed.D.

Lesley Raisor-Becker, Ph.D.
ABSTRACT

Speech-language assessment of bilingual Spanish-English speaking children in the United States has been difficult for researchers and clinicians because there are few assessment measures that have been comprehensively studied with this population and none have shown to be optimal. In recent years, additional assessment measures have been recommended for this population, but few studies have examined the effectiveness of these measures to accurately differentiate children with typical language from children with language impairments. This study examined whether a specific assessment measure is best in differentiating typical language learners and children with language impairment; and whether there are a group of assessment measures that collectively best differentiate the TL and LI groups of children. Selected language assessment measures were administered to 30 Mexican-American 4 and 5 year old bilingual Spanish-English speaking children (15 TL and 15 LI) to determine the predictive value of each assessment measure, as well as which group of predictors can appropriately identify LI and TL children. Using a multiple linear regression analysis, this study found that all of the assessment measures except the dynamic assessment task and NWR English task can be used to assist in the identification of language impairment. Spanish standardized assessments were highly discriminating and strongly correlated with language status. Other assessment measures including the English standardized assessments, language sample, parent interview, and NWR Spanish tasks were highly discriminating and moderately correlated with language status. The results from this study suggest that a parent interview, Spanish and English standardized assessment, and language sample can be administered as a comprehensive battery to identify children with language impairment.
ACKNOWLEDGEMENTS

I have a tremendous amount of gratitude towards my dissertation committee and family for helping me reach this milestone of obtaining my PhD degree in speech-language pathology. There is no way for me to express my gratitude and show my appreciation for all they have done to help make this lifelong dream of mine a reality.

First and foremost, I would like to thank my committee chair, Dr. Nancy Creaghead. She made it possible for me to participate in the PhD program through distance learning. With her leadership and advice, Nancy guided me to continue my studies steadily as a full-time student while working and raising a family. During the past four years, I have learned from Nancy’s knowledge, devotion to the field of speech-language pathology, research views and experiences. I feel blessed and honored to have had Nancy as my academic advisor, mentor and committee chair during my PhD program. She will continue to be a mentor and role model for me. She has taught me to focus on both the forest and the trees, considering the larger picture as well as the details when making decisions. I hope to emulate Nancy’s acceptance of others, flexibility, determination, positivity, and supportive nature and incorporate these traits into my own repertoire of skills.

Dr. Lisa Cahill has assisted me with all the statistical analyses for this study. Her vast knowledge in both speech-language-audiology and statistics made her the perfect researcher to help me with the data analysis. Lisa has always set aside time to assist me even when she was extremely busy. Lisa readily repeated and modified the statistical analyses for my study countless times. I have learned from her patience and gained from her expertise.

Dr. Allison Breit-Smith willingly joined my dissertation committee even though I had not met her previously and she was unfamiliar with my work. While working on this research
project, Allison always offered to help and give advice to better my research project. She guided me with her keen eye for research. She always found a solution/s to any challenge that arose during this research project. She has always been positive and encouraging. I am thankful to her for helping me improve my research skills.

Dr. Jo-Anne Prendeville is an expert on child language and literacy. I have gained immensely as I viewed many of her lectures while teaching and participating in literacy courses for the “Language and Literacy Leadership” project. I appreciate her willingness and effort to include me in mentoring graduate students for the Literacy project using various technologies available. Jo-Anne’s amazing flexibility to accommodate my distance-learning needs and her commitment to joining my dissertation committee right before she retired was remarkable. Her keen understanding of language and research empowered me to ask questions about my project and to think more deeply into the purposes and procedures of research.

Dr. Lesley Raisor-Becker and I go back to the days she and I were teaching a child language course together for Testing and Training International. She encouraged my acceptance into the PhD program and has always recommended opportunities for me. I enjoyed listening to Lesley and gained from her expertise on child language and literacy. She willingly offered to mentor me during an independent study course on statistics. I am also so grateful that she has included me in some of her presentations and on-line courses. I have learned a lot about various teaching practices using on-line formats that she provided. I am amazed at how Lesley is always cheerful and infuses humor in all that she does. She helped me see the light at the end of tunnel when it seemed that I would never reach my goal.

Dr. Carney Sotto recognized that I possessed the skills and desire to complete a PhD degree and encouraged me to reach out to Dr. Nancy Creaghead. She has always looked out for
me to ensure that everything was going smoothly with my program. She has assisted in making a far off dream a reality for me!

Dr. Suzanne Boyce assisted me in submitting two grant applications for this research project. I have learned many techniques about grant writing that I hope to incorporate in the future. I have also learned through her how to critique and analyze research.

On a solemn note, I want to give tribute to Richard Devan for always enhancing the technology so that I could join the classes in real-time and also troubleshooting when there were technological difficulties.

I want to acknowledge Dr. Thomas Campbell, Dr. Kelly Cordero, Dr. Chris Dollaghan, Dr. Kerry Danahy Ebert, Dr. Jocelyne Kalanek, and Dr. Kathryn Kohnert for allowing me to use the non-words they developed for the nonword repetition task. A special thanks for providing me with the actual recorded stimuli so that the task could be replicated for this study. I also want to acknowledge Dr. Maria Adelaida Restrepo for allowing me to use the parental report measure she developed to identify predominantly Spanish-speaking children with language impairment.

I acknowledge Dr. Elizabeth D. Peña for allowing me to use several assessment measures she developed including the dynamic assessment task and portions of the BESA (BESA semantics and syntax subtests). In addition to providing me with these assessment measures, she guided me in selecting the most appropriate measures based on results from her previous studies and experiences.

I thank Dr, Anny P. Castilla-Earls for assisting me in selecting an idea for this research study. Although she was not part of my dissertation committee and affiliated with Suny-Fredonia University, she willingly reviewed several options I was considering for my dissertation and participated in a Skype conference to discuss the feasibility of each one.
I thank Dr. Mahchid Namazi for setting aside time to meet with me about the data analysis for this research study. She reviewed the data and suggested multiple analyses that could be completed with the data collected. Her insights were helpful.

I thank Shira Shabes Wallin for collecting a large amount of the data for this study. Her assistance in collecting the data will help guide future researchers and clinicians in selecting appropriate assessment measures for 4-5 year old bilingual Spanish-English speaking children in the United States.

I thank Adina Weisz for accommodating my work schedule so that I could participate in the courses for my PhD degree. I also appreciate that she arranged for a graduate student to collect the data for this research project. Adina always encouraged me to further my knowledge and professional expertise as a speech-language pathologist. She has supported initiatives to increase collaboration between speech-language pathologists and educators, enhance language/literacy instruction in the classrooms within a response to intervention model, and improve speech-language assessment/intervention practices.

The quintessential thank you goes to my husband, Shimon. He was by my side and provided unwavering support at every step in the PhD program. When there were challenges, he always problem-solved and assured me that “this too would work out” so that I was able to continue along in the program. He often played the roles of both father and mother for our children when I had to meet deadlines. He listened to my ideas, reviewed articles and grants I wrote, and provided suggestions to improve them. Although he is not a speech-language pathologist or researcher, he brainstormed and critiqued with me. His understanding, patience, humor, and intellect made this a shared journey.
I have to thank my children, Avi, Yisroel, Shulamis and Esther. They gave up a lot over the past four years in order for me to complete the PhD and they “rolled with the punches.” They inspire me every day. Without them, none of this would matter.

I thank my parents for always setting a high standard for education and life’s values. They always encouraged me to scale life’s mountains, excel, and reach higher. I thank my in-laws, who have supported this endeavor and showed me how proud they are. I thank my grandparents for their smile of approval and their pride in me for reaching this milestone.

Without the help G-d, I would not have been able to accomplish this milestone. I pray that he will continue guiding me in all my endeavors.
Table of Contents

Abstract ........................................................................................................................................................................2
Acknowledgments ........................................................................................................................................................4
List of Tables ............................................................................................................................................................12
Chapter I: Introduction........................................................................................................................................13
  Statement of Problem ..........................................................................................................................................15
  Purpose of Study ..................................................................................................................................................16
Chapter II: Literature Review .............................................................................................................................17
  Dialect.................................................................................................................................................................17
    Children Speaking a Dialect of English ...........................................................................................................17
    Children Learning English as a Second Language ......................................................................................19
    Consideration of Dialect and Second Language in Assessment ..........................................................21
  Assessment of English-Speaking Children .................................................................................................22
    English-Speaking Children with Language Impairment ...........................................................................24
  Assessment of Bilingual Spanish-English Speakers ...................................................................................26
    Standardized Language Assessments for Spanish Speakers .................................................................27
    Static non-standardized language assessments for Spanish Speakers ..................................................31
      Parental Report .................................................................................................................................31
      Language Sample ..............................................................................................................................33
      Nonword Repetition ..........................................................................................................................36
    Authentic Assessments for Spanish Speakers .......................................................................................38
      Dynamic Assessment ..........................................................................................................................38
    Fidelity .........................................................................................................................................................39
  Spanish-English Bilinguals with Language Impairment ..............................................................................40
Participants’ Age and Sociolinguistic Factors .............................................. 83
Limitations and Future Research ...................................................................... 86
Implications ........................................................................................................... 88
References .............................................................................................................. 90
List of Tables

<table>
<thead>
<tr>
<th>Number</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>54</td>
</tr>
<tr>
<td>2.</td>
<td>56</td>
</tr>
<tr>
<td>3.</td>
<td>59</td>
</tr>
<tr>
<td>4.</td>
<td>60</td>
</tr>
<tr>
<td>5.</td>
<td>62</td>
</tr>
<tr>
<td>6.</td>
<td>65</td>
</tr>
<tr>
<td>7.</td>
<td>66</td>
</tr>
<tr>
<td>8.</td>
<td>67</td>
</tr>
</tbody>
</table>
Chapter I

INTRODUCTION

As the demographics of the United States continue to change toward a more diverse population, it becomes more likely that speech-language pathologists will have students who are culturally and linguistically diverse on their caseloads (American Speech-Language-Hearing Association, 2004). Latinos constitute the largest minority group in the United States and are predicted to be make up a third of the population by 2050 (U.S. Bureau of the Census, 2008).

In order to address the educational needs that diversity brings, policymakers must realize that the under and overrepresentation of minority students in special education often stems from inappropriate assessments used to determine eligibility for special education and related services. Content and linguistic biases, as well as flaws in concurrent validity, are inherent in published language tests for Latino children. This results in inaccurate classification of Latino children as language impaired or typical language learners. It is valuable to examine assessment measures for bilingual Spanish-English speaking children as a step toward developing appropriate assessment measures for all culturally and linguistically diverse students (CLD).

Educators must be aware of the detrimental effect of identifying a CLD child inappropriately. Oftentimes, teachers refer CLD students for special education and related services because of students’ poor performance resulting from limited background knowledge, lack of environmental/linguistic exposure, or demonstration of disruptive behaviors that are inconsistent with mainstream American school culture expectations (Skiba, et al., 2006). In addition, many current assessments used to assess the language skills of CLD children assess children’s linguistic and background knowledge, as well as incorporate assessment practices that are inconsistent with the students’ previous cultural experiences (Roseberry-McKibbin, 1994).
When these children are inappropriately placed in special education programs because of cultural and environmental differences and not a true disability, teacher expectations and students’ performance at school diminish (National Research Council, 2002; (West-Olatunji, Baker, & Brooks, 2006). Inaccurate identification of CLD students for special education and related services also increases the governmental burden of paying for special education services for children who do not present with a language-learning disability and do not require special education and/or related services to access the curriculum. Dollaghan and Horner (2011) state that “accurate diagnosis is a prerequisite to ensuring that costly treatment resources are allocated to all and only those likely to benefit (p. 1077).”

Although federal legislative guidelines advocate for appropriate assessment of bilingual Spanish-English speaking children, currently there is no existing comprehensive, valid, and reliable assessment measure that can identify bilingual Spanish-English speaking children with language impairment. Although standardized assessments have been used as a primary means to identify mainstream American English-speaking children, children who are culturally and linguistically diverse (CLD) may perform poorly on standardized assessments because of unfamiliarity with standardized assessments procedures such as timing of tests and decontextualized format (Roseberry-McKibbin, 1994). They may also be inappropriately identified as language impaired when tests that are designed for native English speakers are used. Even standardized assessments that are modified and translated for a specific language may not be valid or reliable for all minority populations who speak that language. For example, Guttierez-Clellen and Simon-Cereijido (2009) found that Latino children with specific language impairment (SLI) struggle with linguistic forms, such as clitic pronouns (e.g., direct and indirect object pronouns), which is not found among English-speaking children with language
impairment. This is because Spanish requires marking the object pronoun twice as in \textit{Lo vi a el} (Him I saw him) in some contexts and clitic doubling is more commonly used in Spanish than in English (Bedore, 2004). These structures are not typically assessed in formal assessments for English speakers and so a speech-language pathologist might miss a cue signaling a child is at-risk for a language impairment, when using assessments originally developed for English speakers.

Various instruments used to assess the language skills of monolingual and bilingual Spanish speaking children exist; however, none of the measures or combination of measures, currently available has been shown to be adequate for diagnosing language impairment in this population. Dollaghan & Horner (2011) conducted a meta-analysis of the diagnostic accuracy of various standardized and non-standardized assessment measures (e.g., formal word definitions, Spanish Morphosyntax Test, Nonword repetition, parent report of speech and language problems, family history of speech and language problems, mean length of utterance in words, number of errors per t-unit) intended to identify specific or primary language impairment (LI) in United States (US) children who are bilingual Spanish-English speakers and found that there is no optimal published standardized or non-standardized assessment measures yet to be used with bilingual children to classify them as LI (language impairment) or TL (typical language); each has to be supplemented with additional information.

The individual standardized and non-standardized assessments previously developed for bilingual children are beneficial in providing information to the evaluator; however, they cannot identify a language impairment when used alone. Since all of the current standardized and non-standardized assessment measures must be supplemented, it would be beneficial to determine the predictive value of each assessment measure, as well as which group of predictors can best
identify LI and TL bilingual Spanish-English speaking children. The purpose of this study was to determine the assessment measures that best differentiate bilingual typical language learners and children with language impairments, using the current assessment measures that are available and recommended for use with bilingual Spanish-English speaking children. A few assessment measures that have been previously studied were administered. A multiple linear regression statistical analysis was conducted to make accurate predictions about group membership and describe the nature of differences between the groups of LI and TL bilingual children (Warner, 2008, p. 16).

This study examined the following research questions:

1. Which types of assessment [standardized assessment (PLS-4 English, PLS-4 Spanish, CELF – P2 English, CELF-P2 Spanish, Bilingual English Spanish Assessment – BESA – English and Spanish semantics and morphosyntax subtests); parental report (Restrepo (1998) parent interview); language samples (English and Spanish – MLU-w); nonword repetition task (NWR English and Spanish); dynamic assessment (word-learning ability/vocabulary)] best distinguish between bilingual Spanish-English speaking children with language impairment and typically developing children?

2. Is there a group of tests that best distinguishes between bilingual Spanish-English speaking children with language impairment and typically developing children [standardized assessment (PLS-4 English, PLS-4 Spanish, CELF – P2 English, CELF-P2 Spanish, Bilingual English Spanish Assessment – BESA – English and Spanish semantics and morphosyntax subtests); parental report (Restrepo (1998) parent interview); language samples (English and Spanish – MLU-w); nonword repetition task (NWR English and Spanish); dynamic assessment (word-learning ability/vocabulary)]?
Chapter II

LITERATURE REVIEW

Dialect

Children Speaking a Dialect of English

A dialect is a functional and effective variety of a language and a rule-governed linguistic system. Therefore, ASHA’s technical statement on American English Dialects (2003) states that dialectal variations are not considered a speech or language disorder. We need to recognize that regardless of the dialect spoken, an individual may present with a language disorder unrelated to his dialect use. It is easy to mistake a dialectal difference as a speech or language problem. Therefore, speech-language pathologists need to familiarize themselves with the native dialect of the client and recognize whether the errors displayed are due to a dialectal difference or a speech or language problem. In 2003, ASHA’s technical report summarized it in this way:...

...to distinguish between those aspects of linguistic variation that represent regular patterns in the speaker’s dialect and those that represent true disorders in speech and language”.

Consideration of dialectal differences is especially important during assessment of communication disorders to avoid misdiagnosis (ASHA, 2003, p. 2)

When comparing an individual speaker’s dialect to another dialect of the same language, these differences often seem like speech or language problems unless a speech-language pathologist is aware and considers the individual’s native dialect.

Dialect-sensitive language assessments have been developed for children who speak a variation of Standard English such as African-American English (AAE), Southern White English (SWE) and Appalachian English. The purpose of these assessments is to identify children with language impairments, without penalizing them for their dialect use (Ciolli & Seymour, 2004).
Ciolli and Seymour (2004) recommend the use of contrastive items to identify whether “variation due to speech or language dialect” is present and then non-contrastive items to determine whether a language impairment exists. Assessment of subject-verb agreement for the past tense verb *was* is an example of a contrastive item that reliably differentiates between MAE (Mainstream American English) and AAE dialects (e.g., “they was” for AAE and “they were” for MAE). On the other hand, examining whether *was* is deleted is an example of a non-contrastive item that differentiates typical language from language impaired because neither MAE or AAE dialects permit deletion of *was/were* when the subject *they* is used. Pearson (2004) describes an assessment measure for AAE children that was developed to avoid confusing dialect variation with an impairment. The assessment includes stimuli assessing abstract language knowledge that typical children learn and children with LI (language impairment) struggle to acquire. It has shown to reliably distinguish typical development from language impairment, regardless of the dialect used.

Although these studies (Ciolli & Seymour, 2004; Pearson & Ciolli, 2004; Seymour, 2005) support the use of non-contrastive patterns to be used in language tests so that the diagnosis of SLI is more accurate among groups of English speakers, recent research has shown that contrastive patterns should also be used to provide information to assist in the diagnosis of SLI. Oetting and McDonald (2001) examined children with SLI from two non-mainstream dialects, a rural version of Southern African American English (SAAE) and a rural version of southern white English (SWE). Children with SLI produced a few non-mainstream patterns at a higher rate than their age matched peers. These results indicate that dialectal patterns should be used along with non-contrastive patterns to help identify children with SLI who speak non-mainstream patterns of English.
Dialects do not only pertain to English speakers, but exist across all languages. Speakers of any language use a dialect of the language. It follows then that bilingual individuals living in the United States also speak a dialect of a language. They either speak a dialect of a language or of combined languages. Many Latino children in the United States often speak a dialect of their native language of Spanish and learn a dialect of English as the second language. According to Gutierrez-Clellen and Simon-Cereijido (2007a), many Latino children in the United States speak an English-Spanish dialect, which developed due to the use of both Spanish and English spoken in the environment.

Children Learning English as a Second Language

There are many factors that influence a bilingual speaker’s language proficiency including age of exposure to the second language (L2), level of L1 (first language) when L2 was introduced, continued or lack of continued experience with L1, and sociolinguistic variables (Kan & Kohnert, 2008). There is some controversy as to how to differentiate children learning two languages simultaneously and children learning the second language after exposed to the first language for some time (Goldstein, 2004). Children are generally considered simultaneous bilingual language learners when they have been exposed to L2 before the age of 2 or 3. In contrast, the term sequential bilingual language learners, is used to describe children who were exposed to L2 after age 3.

Bilingual children often experience influence from one or both languages (Genesee, 2000). A Spanish-English speaking child may demonstrate usage of some Spanish linguistic features while speaking English and vice versa. The Interactional Dual Systems Model (Paradis, 2001) explains that bilingual children have a separate system for each language that exists early in development; however, some interaction between the linguistic systems may occur. Research
studies on lexicon, syntax and phonology support this model and have shown that bilingual children function with two separate systems, but there is some interdependence between the two languages (Meisel, 1989; Paradis, 2000). According to Gutierrez-Clellen and Simon-Cereijido (2007b), English-Spanish dialects contain influence of Spanish features across linguistic domains (phonology, lexical, morphosyntax). For example, Spanish-English speakers might omit some final consonants. Characteristics of Spanish-English dialects will also differ depending on the variation of English used within the community. For example, features similar to AAE might exist among children in the Northeast region (Gutierrez-Clellen & Simon-Cereijido, 2007a).

Besides for transfer/interference, bilingual children may experience other typical bilingual language learning patterns including code-switching, silent period and language loss. These behaviors are independent of each other and children might demonstrate some of these behaviors when exposed to a second language (Schiff-Myers, 1992). Code-switching is when a bilingual speaker alternates between two distinct language codes (Spanish and English) on the same topic, within and/or across utterances (Kohnert, Yim, Nett, Kan, & Duran, 2005). This is an effective mode of communication for bilingual speakers who share both languages.

When a second language is introduced, some children may not talk much for a while and experience a silent period (Roseberry-McKibbin, 1994). This silent period can last for a few months or for as long as a year and usually goes away on its own. A silent period might occur when a child is first exposed to a second language where he does not talk, but focuses on listening and understanding. It is important to realize that although a child isn’t speaking, he may be listening, understanding and processing the language differences.

Language loss for L1 might occur if the native language (L1) is not continuously maintained and supported. Bilingual children in a subtractive bilingual program model are at a
greater risk for a decrease of language skills and academic success because L1 is not supported (Cummins, 1978; Kohnert, et al., 2005; Winsler, Diaz, Espinosa, & Rodriguez, 1999).

It is also important to be aware that it takes 2-3 years for a bilingual language learner to acquire Basic Interpersonal Communication Skills (BICS) and 5-7 years to acquire Cognitive Academic Language Proficiency (CALP) under ideal conditions (Cummins & Ontario Inst. for Studies in Education, 1981). A child may have developed BICS, but did not yet acquire CALP only because additional time and exposure is necessary for a child to learn and use a second language for academic activities. This is termed the BICS-CALP gap and must also be considered during the assessment of bilingual children because it is a typical second language behavior pattern that might seem like a language delay. In summary, since there are many factors influencing the language skills of bilingual children, dialect, second language learning characteristics and socio-linguistic variables, must be considered during communication assessments with bilingual children.

Consideration of Dialect and Second Language in Assessment

Although we do not yet have published standardized assessments developed for the bilingual population, some progress has been made to develop a test for speakers of various dialects of English (Seymour, 2005). The Diagnostic Evaluation of Language Variation -- Norm Referenced (DELV) (Seymour, Roeper, & deVilliers, 2005) is one test that allows for assessment of children who are speakers of dialects other than Mainstream American English. Although the DELV is a solution to address the problem of children speaking various English dialects, we do not yet have assessment measures that can classify bilingual Spanish-English speaking children with LI and TL accurately. To identify Latino children with language impairments, assessment measures that help distinguish differences due to linguistic transfer from those that are due to a
language impairment are needed, as well as assessment tasks that are culturally appropriate and familiar for the population. Although there are some standardized and non-standardized assessment measures recommended for use with Latino children who speak Spanish and English (e.g., PLS-4 Spanish), these assessment measures cannot be used independently to identify children with a language impairment. As mentioned earlier, we also do not yet know whether a group of assessment measures can better identify Latino children with language impairment.

Assessment of English-Speaking Children

Specific language impairment is a diagnosis used for individuals who only present with a language-learning disability and does not include children with developmental disabilities (e.g., mental retardation, autism, neuromotor impairments) or persistent hearing loss ((Aram, Morris, & Hall, 1993). Paul (2007) states that a speech-language evaluation for a mainstream American English speaking child should include standardized assessments to determine whether the child’s performance is significantly deviant from typically developing children among other assessment methods. Published standardized language assessment measures exist for use with mainstream American English-speaking children to accomplish this task. There is some controversy as to the cutoff score that should be used to identify children with SLI (specific language impairment). Tomblin, et al. (1997) state that there is no consensus of a reference standard for the identification of mainstream American English speaking children with LI. Some researchers recommend a cutoff of 2 standard deviations from the mean (Bloom & Lahey, 1978) and others recommend the 10th percentile (Paul, 2007). Although degree of delay for inclusionary criteria of SLI varies, standardized assessments for mainstream American English-speaking children are used to differentiate language impairment from typical language ability. Spaulding, Plante, and Farinella (2006) examined 43 standardized tests to determine whether these tests identify
children with language impairment, using the standard cutoff score recommended in the respective studies examined. The evidence did not support their use in identifying language impairment in children and many of the children performed above 1.5 or within 1 standard deviation of the mean even though these children had a language impairment. There were only 5 tests [(the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF–4; Semel, Wiig & Secord, 2003), Preschool Language Scales—Fourth Edition (PLS–4; Zimmerman, Steiner, & Pond, 2002), TEGI (Test of Early Grammatical Impairment; Rice & Wexler, 2001), Test of Language Competence—Expanded Edition (TLC–E; Wiig & Secord, 1989), and Test of Narrative Language (TNL; Gillam & Pearson, 2004)] that actually met the sensitivity and specificity scores to accurately identify mainstream American English speaking children with LI.

Standardized assessments only include a few items assessing specific skills; thus, they cannot be used to determine particular areas of deficit and plan intervention program/s and target/s. Non-standardized assessment measures (criterion-referenced assessments, authentic assessments) must be supplemented to examine the particular areas of deficit because standardized assessments do not provide this detailed information (Paul, 2007). Language sampling is an example of a non-standardized assessment that can be used to describe a child’s linguistic strengths and weaknesses to assist with intervention planning. Although this method is appropriate and effective for mainstream American English-speaking children, this method might not be effective with culturally and linguistically diverse children because of the problems that exist in translated and adapted versions of standardized assessments and the socio-linguistic variables that must be accounted for as well. In addition, Latino children in the United States are exposed to Spanish and English resulting in different language skills and proficiency levels in each language.
Authentic assessment is, “a process of gathering information through which the skills and needs of a student are identified with respect to the language and curricular demands they will encounter (Prendeville & Wellman, 2011).” There are many different assessment tasks that may be conducted during an “authentic assessment” including observations, ethnographic interviewing, structured probes, behavioral sampling, curriculum-based assessment, dynamic assessment, checklists/rating scales/rubrics, classroom performance communication domains, portfolios, and language sampling in multiple contexts. In an authentic assessment, a criterion is set for the minimum level of communication competence a child must perform in a real-life learning situation, in or out of the classroom (Schraeder, Quinn, Stockman, & Miller, 1999). Schraeder, et al. (1999) examined an authentic assessment tool and found it to be valuable for screening preschool children from multicultural backgrounds. There has been an increase in the use of authentic assessments to evaluate speech and language skills because of increased awareness regarding additional variables (e.g., interactional partner, setting, task, materials) that influence performance. These variables can be assessed in a naturalistic learning environment, but cannot be administered in a standardized format.

*English-Speaking Children with Language Impairment*

Numerous research studies examined the clinical markers for language impairment in English speaking children. Young children with SLI have novel word-learning difficulties, where they struggle to learn to comprehend and produce new words, resulting in smaller initial vocabularies and learning words at a slower pace than typical language learners (Gray, 2005). Gray (2003) compared the fast-mapping and word-learning abilities between typical language learners and children with SLI. The findings from this study indicate that typical language learners learned words more quickly, as well as comprehended and produced more words than
the SLI group. Children with SLI require additional exposures to new words in learning them (Gray, 2003). The research shows that children with SLI have difficulty fast-mapping and learning new words (Dollaghan, 1987; Gray, 2003)

Children with SLI also produce fewer grammatical morphemes than typically developing children, matched for mean length of utterance (Leonard, et al., 2007; Rice & Wexler, 1996). On average, TL children master the use of grammatical morphemes by age 5 (Rice & Wexler, 1996); however, children with SLI continue to make errors on the production of grammatical morphemes, with some errors persisting through adolescence. Specifically, children with SLI have difficulty with finite verb morphology. Moyle, Karasinski, Weismer, and Gorman (2011) examined three variables, which can possibly discriminate TD and SLI children; finite verb morphology, noun morpheme and mean length of utterance in morphemes. Analysis of several language transcripts from 50 typically developing children and 50 children with SLI showed that only a combination of the three grammatical morpheme composites adequately differentiated children in the two groups. Although there is strong research suggesting that verb morphology (e.g., past tense –ed, third person singular –s) is a clinical marker to identify SLI, Moyle, et al. (2011) indicate that a verb morphology composite cannot be used alone and must be combined with other measures.

Besides for verb morphology and noun morpheme accuracy, NWR performance might also be considered as a co-clinical marker in the identification of SLI. Children with SLI present with a selective impairment in repeating lengthy non-words and their errors increase as the number of syllables in the non-word increases. Although children with SLI are frequently identified based on their language deficits, these children are also impaired in their NWR performance (Gathercole, 2006). The use of NWR as a clinical marker is reasonable because of
the association between NWR and SLI (Gathercole, 2006). However, since NWR is only
associated with SLI and does not cause SLI, NWR performance should be used along with other
clinical markers in the identification of SLI and TL groups of children.

In addition, a few research studies have examined clinical markers for language
impairment among minority English-speakers, specifically African American English speakers.
Unique clinical markers for minority English-speaking populations are necessary because some
AAE (African-American English) features resemble the speech-language patterns of delayed
speakers of SAE (Standard American English) (Oetting & McDonald, 2001). A recent study by
Garrity and Oetting (2010) investigated Be auxiliary verb forms (am, is, are) among AAE
children with SLI, age-matched and language matched controls. Overall, children in the SLI
group consistently produced the BE form less frequently than the other groups. Leonard, Miller,
and Gerber (1999) supported the use of a finite verb morphology composite to identify
mainstream American children with language impairment because children with SLI have
significant difficulty with these morphemes. However, a verb morphology composite cannot be
used with AAE children because even typically developing AAE language-learners will use
different forms for past tense –ed, third person singular –s, and copula and auxiliary be forms
(verb morphology composite) compared to MAE (mainstream American English) (Pearson,
2004).

Assessment of Bilingual Spanish-English Speakers

The following assessment measures have been suggested for use with the bilingual
population thus far; ethnographic interviewing (Westby, Burda, & Mehta, 2003), parental report
(Restrepo, 1998; Thal, Jackson-Maldonado, & Acosta, 2000), standardized assessments that are
translated and/or adapted for the child’s population, language sampling (Gutierrez-Clellen,
Restrepo, Bedore, Pena, & Anderson, 2000), dynamic assessment (Laing & Kamhi, 2003) and non-word repetition tasks (Campbell, Dollaghan, Needleman, & Janosky, 1997). Each of these assessment measures has shown positive statistical results in helping to classify children with LI vs. TL children and reduces the number of children mistakenly identified as having a language impairment.

It is important to recognize that all of the assessment measures listed have been developed and used for mainstream English speaking children. In recent years, research studies have examined the effectiveness of these assessment measures with children from culturally and linguistically diverse backgrounds. Due to varied previous knowledge and experiences, cultural and linguistic differences, sociolinguistic variables, and socio-economic status, it has been difficult for researchers and clinicians to select appropriate assessment measures for culturally and linguistically diverse populations in the US. Therefore, many researchers have studied existing and modified assessment measures in an attempt to determine assessment measures most appropriate for culturally and linguistically diverse children. Since Latinos are the largest minority group in the United States (U.S. Bureau of the Census, 2008), there is an urgency to determine appropriate standardized and/or non-standardized assessment measures that can be used among the various cultural and linguistic groups of Spanish-speaking children in the United States. In the United States, Latino children represent various linguistic and cultural groups resulting in various discourse styles, Spanish dialects, and levels of bilingualism (McCabe & Bliss, 2003). We should attempt to develop assessment measures that can be used for Latino children from various linguistic and cultural groups.

*Standardized language assessments for Spanish-English speakers*
Traditional standardized assessment is a static measure given at one time and the scores are based on national norms. These tests are administered using similar procedures as standardized assessments for mainstream American English speakers, where the clinician administers the test during a testing session in a decontextualized format, adhering to the standardized procedures delineated within each specific test.

Published standardized assessments currently available for Spanish (and Spanish-English) speakers are based on cultural, academic and social expectations for middle class English speaking children in the United States. These tests are problematic for Spanish-speaking children because of the content and linguistic biases that exist, different language socialization practices between Spanish speakers and middle class English speakers, and significant validity problems inherent in them. Some content and linguistic biases have been reduced in assessments for bilingual Spanish-English children by replacing some of the pictures and language stimuli to be consistent with language socialization practices for Spanish speakers. Although some biases were minimized, it is impossible to remove all biases from a standardized assessment developed for a different cultural/linguistic group, specifically mainstream American English speakers. For example, the PLS-4 assesses a child’s ability to identify body parts and clothing items; however, this is not taught in some Spanish-speaking communities during language socialization with their children (Gutierrez-Clellen, Restrepo, & Simon-Cereijido, 2006). In the PLS-4 and CELF-P2 Spanish Editions, item progression was modified from the English version to match the developmental trend for the Spanish-speaking population. Several standardized tests have also been translated and adapted from the English version. These tests contain some items that are useful in the identification of English speakers, but are not useful in the classification of Spanish-speakers. Bedore and Leonard (2001) found that the inclusion of
possessive pronouns, third person singular verbs, and prepositions in Spanish assessments such as the PLS-4 Spanish and CELF-P2 Spanish did not differ among SLI and control groups for Spanish speakers in the United States. Bedore and Leonard (2005) also found that Spanish speakers with language impairment had difficulty with definite articles and direct object clitics, which differs from English-speaking children with language impairment and so these linguistic skills are not included in tests initially developed for English-speaking children. English tests also often include items on finite verb morphology; however, the current research shows that Spanish speakers produce present and past (preterite) tenses for verbs accurately (Bedore & Leonard, 2005). This might be the reason that these tests fail to achieve the sensitivity and specificity values set by Vance and Plante (1994); thus, they do not adequately classify LI and TL children.

These test modifications have only decreased the mean distribution. Latino (and CLD) children still perform below the mean but within normal age limits so that fewer children are misclassified (Laing & Kamhi, 2003). Also, since bilingual children vary in their levels of proficiency, standardized assessments appropriate for this population must collect a normative sample of many children with varied levels of proficiency (Goldstein, 2006). Levels of proficiency for each language varies depending on the child’s continued experience with L1, age of exposure to L2, level of L1 when L2 was introduced, amount of input for each language, aspect of language being measured and socio-linguistic variables (Kan & Kohnert, 2008). “Since bilingual children are not equally proficient in both languages, and may have different proficiencies in different parameters and modalities, standardized measures normed on monolingual children cannot provide a valid picture of the ‘underlying’ language abilities of bilingual children (Goldstein, 2006, p. 313).” Therefore, bilingual assessments should occur in
both languages the child speaks. Significant construct validity issues are also still present in these tests due to an inappropriate reference standard when identifying the sensitivity and specificity groups for the discriminant accuracy analysis. These tests still do not meet the standards for concurrent validity established by Plante and Vance (1994). At this time, published tests for Spanish-speaking children in the US do not meet the 80% sensitivity standard set by Vance and Plante (1994), making each test on its own inappropriate for diagnostic purposes and their specificity rates frequently results in the overrepresentation of typically developing Spanish and Spanish-English speaking children (Gutierrez-Clellen, et al., 2006).

Although many researchers have criticized the use of standardized assessments for minority students because they were not developed specifically for bilingual children, a normative sample was not collected for the bilingual population tested, and/or they do not have acceptable concurrent validity; standardized assessments provide the evaluator with some information about the language skills of the child being evaluated. With consideration of the inherent problems, standardized assessments suggested for use with predominantly Spanish and bilingual Spanish-English speaking children help the researcher and clinician somewhat to classify children as LI or TL. For example, some standardized tests such as the PLS-4, CELF-4 and CELF-P2 Spanish Editions have been modified to reduce some of the content and linguistic biases, as well as collect a separate normative sample.

Since bilingual Spanish-English speaking children learn English as a second language, clinicians and researchers frequently complete some of the language assessment measures in English. Interpretation of results from English tests should be interpreted cautiously because a study by Paradis (2005) indicates that typical English Language Learner (ELL) children made similar morphological errors as English-speaking children with LI on the Test of Early
Grammatical Impairment (TEGI; M. Rice & K. Wexler, 2001). The typical ELL children scored in the clinical range on these English morphological tasks even though language impairment was not suspected in the group of children studied (Paradis, 2005). The data in this study was collected from 24 typically developing language minority students who received English as a Second Language (ESL) support services for several months.

It should be noted that a standardized assessment is in the process of being developed, designed and tested for Latino children in the United States. The Bilingual English Spanish Assessment (Pena, in development) includes separate Spanish and English measures, which were developed based on research related to language development and cultural issues for Spanish and English speakers respectively. This test assesses various linguistic domains including semantics, syntax, pragmatics and phonology.

Given the scarcity of standardized assessments specifically developed for bilingual children, researchers have suggested that alternative assessments be used (Rojas, 2009). It should be noted that even if problems with standardized assessments did not exist for this population; one assessment measure would not be sufficient to determine the presence of a language impairment. Every language disorder diagnosis must be made using multiple assessment measures regardless of the population tested (Goldstein, 2006; Paul, 2007).

*Static non-standardized language assessments for Spanish speakers*

**Parental report**

Parental report measures are a means for the clinician/researcher to obtain information regarding a child’s speech-language development from the parents’ perspective. These are usually self-rating scales. Parents are typically asked to complete a checklist, observe and rate their child’s behavior and/or respond to questions about the child’s language skills. These
measures are completed by the parent and the clinician might also be present to explain some of the items included. They are easily administered, cost-effective, and sometimes more representative than a spontaneous language sample collected at one point in time in a clinic, school, or home setting (Dale, 1991). Research on mainstream American English speaking children, as well CLD children, have shown that parental report is a valuable measure to assist in the identification of speech-language delays. Studies have shown that parents are reliable informants and can provide accurate descriptions of their child’s language development and skills; however, there is still hesitancy in using parental report as the only or primary means of assessment (Dale, 1991). This is because parents are not trained to rate their children’s linguistic forms and may be biased towards their own children’s language abilities; thus, they may not provide information accurately. Fenson et. al. (1993) cautions clinicians about the need to use additional measures besides parental report, especially when making decisions about intervention. Most studies used the Macarthur Developmental Inventory in Spanish and/or English as a parent report measure and examined children’s vocabulary and syntax development, which proved to be a valid measure (Dale, 1991; Thal, et al., 2000; Thal, O’Hanlon, Clemmons, & Fralin, 1999). Other researchers have created their own parental report measures such as the parental report measure by Restrepo (1998) and a parent interview section included in the Bilingual English Spanish Assessment (BESA) by Pena (in development). In contrast to the Macarthur Developmental Inventory where parents note specific vocabulary words and syntactic structures they have heard their child produce, the BESA (Pena, in development) and Restrepo (1998) parental report measures collect information based on parents’ interpretation of their child’s overall speech-language skills. The BESA also collects information about the number of years and amount of exposure in each language for bilingual children. In this study, the parental
report measure by Restrepo (1998) was used because its accuracy to distinguish predominantly Spanish-speaking children with language impairment and typical language learners is close to the minimum set by Plante and Vance (1994).

Ethnographic interviewing is a means of asking the right questions during an interview that allows the clinician/researcher to understand the world from a client’s perspective. It is a culturally sensitive technique to obtain information regarding a child’s culture through observations and interviews (Westby, 2003). Ethnographic Interviews seek an understanding of the family’s culture from the parents’ perspective or from a member of that culture (Westby, 2003). This method maximizes the collection of accurate information from parents with different cultural practices. There are several guiding principles to consider when conducting an ethnographic interview including exploring everyday routines across social dimensions, following parent’s lead in conversing about their child, using open-ended questions to allow parents to tell their story, using contingent open-ended questions to obtain follow-up information and responding with relevant comments to validate and support the families’ views (Brown & Woods, 2010). Given that this research study will examine language skills in Latino children, previously developed parent interviews will be used, which already incorporate culturally sensitive questioning techniques. Since parental report measures and interviews only provide a window into a child’s developing language skills and cannot on its own identify a child with a language impairment, other sampling or assessment procedures were supplemented (Maldonado & Conboy, 2007).

Language Sample

Language sampling assesses spontaneous speech-language skills and allows the child to display his/her language knowledge in a naturalistic context. The clinician or researcher collects
a language sample in a naturalistic context (e.g., play, storytelling) and then analyzes the sample to determine a child’s speech-language skills.

Language sampling is a reliable measure when analyzing the speech-language skills of students who are culturally and linguistically diverse (Gutierrez-Clellen & Simon-Cereijido, 2009). Gutierrez-Clellen and Simon-Cereijido (2009) specifically studied the use of language sampling with predominantly Spanish-speaking and bilingual Spanish-English children by collecting a language sample in both languages, which provides more information regarding the speech-language skills of participants than a language sample collected in only one language. In the present study, the research team attempted to elicit both Spanish and English language samples from the participants; however, there were many students who only retold the story in one language. Although bilingual children with a language impairment are usually defined as having lower language skills in both L1 and L2 than their age- and experience-matched peers (Association, 2004; Kohnert, 2008), many young bilingual children in the United States either experience language loss in L1 or do not yet have sufficient exposure to acquire L2; thus, their language skills in one language might be poor and/or cannot be assessed in at least some areas. When a situation like this arises, a clinician would ideally obtain a language sample in both languages and then use the child’s better (more complex) language sample to determine the presence or absence of a language impairment. In a study by Windsor, Kohnert, Lobitz, and Pham (2010), children only needed to score within the normal range on either the Spanish or English CELF-4 subtests administered in order to be included in the typical bilingual group. In a different study by Bedore, Pena, Gillam, and Ho (2010), a language composite score was derived by using the children’s higher language performance on each of the English/Spanish subtests administered (all of the subtests have English/Spanish versions) to determine the best predictors
of language ability. The authors explain that this was done in order to ensure that the child’s language ability was examined and not language proficiency. A similar procedure was implemented in this research study where only the participants’ better language sample was included in the data analysis in order to obtain the best estimate of the child’s language abilities.

Gutierrez-Clellen and Simon-Cereijido (2009) suggest calculating an MLU for morphemes using an English narrative language sample and an MLU for words with the Spanish narrative language sample, which both can be calculated using the SALT software. However, since in this study only the participants’ better language sample was included in the data analysis, cross-linguistic comparisons were made in this study and Gutierrez-Clellen, et al. (2000) then suggests that MLU-w is the most accurate procedure in this scenario to be used for a gross measure of sentence length. These authors also state that MLU-w is the best measure when code-switching occurs during the language sample and many of the participants code-switched during production of the language samples.

Elicitation techniques that will foster productivity and complexity when collecting a language sample also need to be considered (Restrepo & Castilla, 2007). There has been some research indicating that story retelling tasks foster more complex language and seems to be the most sensitive procedure to identify a language impairment in Latino children (Restrepo & Castilla, 2007). Restrepo & Castilla (2007) reported a previous study they completed with Mexican-American typical and atypical language learners on three different elicitation techniques including interview (adult led conversation), picture description and story retelling. The results from this study indicated that story retelling tasks resulted in more errors among children with atypical language skills, suggesting that it is a sensitive measure in detecting language problems. It should also be noted that although there is scarce research on narratives
and elicitation techniques for Latino children, it seems that children from this population primarily talk about families more than any other topic within their narratives (McCabe & Bliss, 2003).

Although language sampling is beneficial for use with Latino children, it is a static measure and takes a significant amount of time to administer (Stockman, 2010). In addition, there have been a few research studies indicating that a child’s cultural and linguistic norms will influence familiarity and performance on a task so that some Latino children will not perform well on a language sample or standardized assessment due to lack of linguistic and/or environmental exposure. Many children from Spanish-speaking backgrounds in the United States are socialized differently in their home culture, which leads to poor performance on mainstream American language tasks. Pena, Bedore, and Rappazzo (2003) investigated Spanish, English and bilingual children’s performance on a battery of semantic tasks and found that bilingual Spanish-speaking children scored higher on expressive and receptive Spanish object-function items (e.g., show me what you do with a scissor). This may be because emphasizing the names for objects is less representative of the language input in Spanish learning children; however, it is emphasized by parents of English learning children. Also, Latino (Puerto Rican) children and African-American children performed better on an item description task than a labeling task (Pena & Quinn, 1997). Since task familiarity is influenced by cultural experience, and one group might be at an advantage on a given task because of task familiarity rather than language ability, processing dependent measures might provide more accurate information regarding a child’s language learning ability.

*Nonword Repetition*
Non-word repetition (NWR) tasks are a processing dependent measure, where individuals repeat non-words. An individual is asked to repeat non-words, increasing in syllable length, immediately after the word is presented auditorally. Accurate repetition of non-words is influenced by word-likeness (Edwards, Beckman, & Munson, 2004), phonotactic probability (Edwards, Beckman & Munson, 2004), articulatory complexity (Archibald & Gathercole, 2006), prosodic patterns (Roy and Chiat, 2003) and non-word duration (Gathercole, Archibald & Joanisse, 2009).

A few studies examining NWR with culturally and linguistically diverse students indicated that it is a promising task for use with culturally and linguistically diverse students. Campbell, et al. (1997) explain that NWR tasks are free of culture and income bias. Washington and Craig (2004) used a NWR task with a group of AA (African-American) children as one of the measures in a screening, which proved to be predictive of a language impairment, along with some other valid assessment measures for this population including the Kaufman Nonverbal Scale (KNVS; Kaufman & Kaufman, 1983) and NDW (number of different words) calculated using a 20-minute spontaneous language sample during free play. Gutierrez-Clellen and Simon-Cereijido (2010) and Windsor, et al. (2010) indicate that NWR tasks need to be given in both languages for bilingual children because varying levels of Spanish and English proficiency influence NWR performance. Currently, there isn’t sufficient data to use previously developed NWR tasks as a diagnostic marker for bilingual Spanish-English children in isolation; however, it might be helpful in predicting a LI along with other variables. In the Windsor, et al. (2010), four groups of children (TD monolingual English, TD sequential bilingual Spanish-English, LI monolingual English, LI sequential bilingual Spanish-English) completed NWR tasks in Spanish and English to examine the usefulness of these tasks as a clinical marker for SLI. The outcome
indicated that these tasks cannot act as a sole clinical marker for SLI, but might be a useful composite marker. A cutoff score of 78% for the English NWR task and 80% for the Spanish NWR tasks were used, which were determined to have adequate clinical power in differentiating typical developing and language impaired groups.

**Authentic Assessments for Spanish Speakers**

*Dynamic Assessment*

“Dynamic assessment is a process-oriented approach to evaluation that looks at children’s responses to learning situations rather than at the traditional assessment measures that examine only product-oriented, static measures of performance (Ukrainetz, Harpell, Walsh, & Coyle, 2000, p. 142).” Dynamic assessment is an approach where the speech-language pathologist monitors the child’s learning of a specific set of behaviors within the context of scaffolding as opposed to static measures that assess a child’s current level of functioning (Laing & Kamhi, 2003). Some types of dynamic assessment are test-teach-retest (Pena, Iglesias & Lidz, 2001), object/activity manipulation (Pena & Quinn, 1997), and graduated prompting (Larsen & Nippold, 2007).

The dynamic assessment process has been used extensively to differentiate children from culturally and linguistically diverse backgrounds with typical language ability and children with language impairment (Kapantzoglou, Restrepo, & Thompson, 2011; Pena, Iglesias, & Lidz, 2001; Pena & Quinn, 1997). This is because bilingual language learners’ dominance and proficiency varies with the amount of experience with each language throughout the language acquisition process making it difficult to use static measures (Kohnert & Bates, 2002), which assess existing language knowledge (Kapantzoglou, et al., 2011). Therefore, it seems necessary
to include dynamic assessment tasks when conducting assessments to make differential diagnoses for this population.

Research studies have proven that dynamic assessment is successful with Latino children because it does not assess previous knowledge, but allows the child to demonstrate what they are able to learn (Gutierrez-Clellen & Pena, 2001). It helps to identify the language-learning potential of the child and differentiate between a language difference and impairment. Based on a review of the different dynamic assessment methods, Gutierrez-Clellen and Pena (2001) recommend the use of a test-teach-retest method to assist in differentiating between a language impairment and a language difference.

In this research project, a word-learning dynamic assessment task was used. A narrative dynamic assessment task as described by Pena, et al. (2006) was considered as well; however, narrative skills are only emerging in some 4-5 year old children and not yet acquired at this age to differentiate TL and LI children.

Fidelity to the Study Procedures

An important component to any research project is adherence to the proposed procedures while the actual study is conducted. Therefore, it is necessary to consider the issue of fidelity, which is the degree that the administration procedures and intervention programs are implemented as planned (Mowbray, Holter, Teague, & Bybee, 2003). In this research project, it was crucial that administration of assessment measures occur as planned for all participants so that accurate outcomes are obtained. Outcome results might vary depending on the competence of the researchers conducting the assessment measures during data collection (Mowbray, et al., 2003). For example, an experienced bilingual speech-language pathologist might administer the
assessments more skillfully, resulting in better performance and outcomes than a novice speech-language pathologist or research assistant.

Fidelity criteria needed to be established, data collected and the indicators needed to be examined to determine validity and reliability. To establish fidelity criteria, a meta-analysis completed by Dollaghan and Horner (2011), as well as a literature review of assessment measures for bilingual Spanish-English speaking children completed by the primary researcher, indicated that the several assessment measures included in this study were documented as helpful for making the LI or TL diagnosis with this population.

The results from a research study by Sterling-Turner, Watson, and Moore (2002) indicated that direct training can increase fidelity of treatment program implementation compared to indirect training. Performance feedback has also shown to improve implementation accuracy (Burns, Peters, & Noell, 2008; Noell, Witt, Gilbertson, Ranier, & Freeland, 1997), thus, increasing fidelity of the administration process.

Spanish-English Bilinguals with Language Impairment

The learning of two languages does not place a child at-risk and/or cause a language impairment. Paradis, Crago, Genesee, and Rice (2003) examined the similarities and differences among monolingual French-speaking children with SLI, monolingual English-speaking children with SLI and bilingual French-English speaking children with SLI. The authors found that bilingual and monolingual children with SLI have the same tense-marking morphology deficits, typical for SLI. These results suggest that bilingualism might not interfere with overall language acquisition and/or exacerbate a language impairment.

Cross-linguistic research (e.g., Germanic, Romance languages) on SLI indicates that children with SLI have significant difficulties with grammatical skills; however, the specific
grammatical morpheme types that are problematic differ across languages (Paradis, et al., 2003). A description of the grammatical morphemes problematic for children with SLI in each language is necessary before identifying children who speak foreign languages with language impairment. A few research studies with monolingual Spanish-speaking children with SLI reveal linguistic deficits of the production of morpho-syntax including articles, direct objects and verb inflection.

Spanish-speaking children with SLI show difficulties with the production of articles during spontaneous samples and elicited probing tasks. Some studies have found that article omissions were the most common article error types (Anderson, Souto, Iacute, & A, 2005; Bedore & Leonard, 2001) and others have found that gender agreement errors were more frequent among Spanish speakers with SLI than omissions (Restrepo, Iacute, A, & Gutierrez-Clellen, 2001). Spanish requires a clitic pronoun to denote direct objects. Studies with preschool children on the production of object clitic pronouns in a probing task revealed that children with SLI have difficulty with production of this form and omissions was the most frequent error pattern, along with some gender errors (Bedore & Leonard, 2001; Jacobson & Schwartz, 2002). The few Spanish verbal inflections (person agreement – first, third; tenses – perfect preterite, present indicative) that were examined indicate that verb inflections are not problematic Spanish-speaking children with SLI (Bedore & Leonard, 2001).

In addition to describing SLI characteristics in Spanish and other languages, it is important to consider variables specific to bilingual children and the impact on the determination of SLI among bilinguals. First, typical bilingual language learning patterns might appear as a linguistic problem. For example, it is sometimes difficult to tease out whether the grammatical errors are due to “language loss” or a language-learning disability in acquiring L1. Restrepo and Kruth (2000) completed a longitudinal case study of a bilingual Spanish-English speaking child,
learning English as a second language, who presented with SLI. In this study, the participant was experiencing language loss and displayed limited use of verb tenses as compared to a typical peer. This participant’s situation is similar to many bilingual Spanish-English second language learners in the US experiencing a language contact situation where English is used more frequently than Spanish. Even though current research suggests that verbal inflections is not a significant difficulty for Spanish speaking children with SLI, bilingual second language learners with SLI might use a limited set of verb tenses due to limited interaction with Spanish-speaking adults and language loss of L1.

Second, there are some communication patterns that are similar among monolingual children with SLI and bilingual children learning a second language (L2). Paradis and Crago (2000) examined seven year old children’s production of morpho-syntax forms (tense morphology, temporal adverbials, agreement morphology, and distributional contingencies associated with finiteness), using spontaneous language samples from French-speaking children with SLI, English-speaking L2 learners of French, and French-speaking controls. The results indicate that monolingual children with SLI and bilingual L2 children showed similar production of morpho-syntax forms and errors. In this study, the monolingual children with SLI and bilingual L2 language-learners primarily produced more non-finite verb forms than the typical monolingual age-matched controls. The SLI and bilingual L2 children also showed less accurate production of the past and future verb tenses than the age-matched controls. Although the Paradis and Crago (2000) study investigated bilingual second language learners, simultaneous bilingual language learners might also experience similar communication patterns as monolinguals due to influence and code-switching between languages. Also, these children rarely have equal continuous exposure to both languages.
Since bilingual children experience typical second language learning characteristics that sometimes appear similar to English-speaking children with language impairment, it is a challenge to describe characteristics of language impairment in bilingual children. A review of previous research on articles with Spanish speaking children reveals inconsistency with the article error types among Spanish-speaking children with SLI. In the studies where participants with SLI were exposed to Spanish as the majority language, omissions were observed. In the studies where participants were in a language contact situation where English was the majority language, gender agreement errors frequently occurred among children with SLI. Anderson (2007) provides a reasonable explanation that Spanish-speaking children in the United States have limited exposure to Spanish and formal schooling in English, which creates problems for these children with noun phrase (NP) gender-agreement, resulting in frequent gender-agreement errors.

Besides for grammatical deficits, young Spanish-speaking children with language impairment have word-learning difficulties. A recent study by Kapantzoglou, et al. (2011) examined the usefulness of a dynamic assessment method to identify Spanish-English speaking children with language impairment, which proved to be promising for use with bilingual children. The findings also indicated that children with SLI are slower in making associations between semantic and phonological properties with the targeted word. These results are consistent with the findings from English-speaking children with language impairment.

Identifiers of Spanish and Spanish-English Speaking Children with Language Impairment

Recent studies on bilingual assessment (Gutierrez-Clellen, et al., 2006; Gutierrez-Clellen & Simon-Cereijido, 2007a; Restrepo, 1998; Restrepo, et al., 2010) have examined Latino children’s performance on isolated language tasks discussed in previous sections and the
accuracy of those tasks in identifying children with language impairment. For example, Restrepo (2010) showed that a grammatical measure, the S-MST, was able to classify 4-6 year old Spanish-speaking children with language impairment (LI) with good accuracy (fair sensitivity and good specificity). There has been only one study that examined whether a collection of assessment measures can classify predominantly Spanish-speaking children with language impairment accurately. Through a discriminant analysis, Restrepo (1998) showed that a combination of two measures, parental report and number of errors per t-unit, can be used to identify Spanish-speaking children with language impairment. Together, these two assessment measures had a sensitivity value of 93.7% and specificity value of 100%, which exceed the criterion of 80% established by Plante and Vance (1994). Over the past ten years, additional assessment measures were developed and tested for use with bilingual Spanish-English speaking children. However, there have not been recent studies that compared current assessment measures recommended for use with Latino children to determine which assessment measures are most useful in diagnosing a language impairment for children in this population.

**Terminology**

The terms predominantly Spanish speaking, predominantly English speaking and bilingual Spanish-English speaking are used in the literature when referring to bilingual Latino children (Goldstein, Fabiano, & Washington, 2005; Hammer, et al., 2012; Kapantzoglou, Restrepo, & Thompson, 2012; Pena, et al., 2003; Restrepo, 1998). The specification of “predominant” is often used when language dominance was measured through parent and teacher surveys and proficiency ratings, as well as participant language assessments (e.g., vocabulary). There is considerable debate as to whether it is possible to determine language proficiency because levels of proficiency vary between tasks and contextual support and there are sometimes
only small differences present between the languages making those differences difficult to measure. It might be impossible to make an empirical claim that a bilingual speaker is equal, more, or less proficient in one of his languages than the other (Hulstijn, 2012). So, in this research project, the term Spanish-English speaking children was used consistently as it is more conservative.

In reference to the term language impairment, primary language impairment and specific language impairment were used to describe a child with a language impairment without significant hearing loss, medical problems, or cognitive problems (neurological, developmental).
Chapter III

METHODS

Subjects

Thirty bilingual Spanish-English speaking children (15 children with typical language [TL] and 15 children with language impairment [LI]) from low socio-economic status (SES) families participated in this study. SES was assessed based on children’s eligibility for free or reduced-price lunch. Twenty of the participants were males (7 TL and 13 LI) and 10 were females (8 TL and 2 LI). All children were recruited from a public school program in a metropolitan area in the Northeast. The district is comprised of a large culturally and linguistically diverse population. Of the approximately 5,000 students enrolled, over 4,000 receive free or reduced lunch. Seventy percent (70%) of the students speak Spanish as their home language. Children who served as subjects attended full-day preschool or kindergarten programs with most of the curriculum instruction in English. All participants met the following criteria:

(1) working-class socioeconomic backgrounds,

(2) Mexican-American descent,

(3) Spanish-speaking families (caregivers reported Spanish to be their primary or sole language),

(4) began the systematic learning of English as a second language in school in the United States, between 3 and 5 years of age although all participants resided in the United States so that exposure to English was present prior to beginning formal education (e.g., via television, older siblings, community professionals, etc.).

(5) no significant hearing loss, medical, or cognitive problems (neurological, developmental).
Since no widely agreed formal reference standard exists for language impairment among Latino children, prior enrollment in treatment of 4 months was the reference standard for the children with LI. According to Dollaghan and Campbell (1998), prior treatment can provide a reasonable starting point for investigation, especially if supported by evidence of their validity and reliability. In addition, Lewis (1995) supports the use of clinical judgment as a reliable measure to identify a disorder and Frontera and Horowitz (1995) showed that teachers have been found to assist in the identification of children at-risk for language and learning disabilities accurately. Restrepo, et al. (2001) used clinical judgment by an ASHA certified bilingual Spanish-English speech language pathologist with at least 3 years of experience and a teacher questionnaire to identify Spanish speaking children with SLI (specific language impairment) because no one measure is optimal and can be used alone. Since prior treatment cannot be used as an only measure in the identification of language impairments and there are no other gold standard measures, converging evidence using prior treatment, (Dollaghan and Campbell (1998), teacher report (Frontera and Horowitz (1995), speech-language evaluation by a bilingual Spanish-English speech-language pathologist with at least 3 years of experience (Restrepo, et al. (2001), and clinical judgment (Lewis, 1995) were used to identify children in TL and LI groups. All children with TL met the following criteria: (1) enrolled in a preschool and/or kindergarten classroom environment for at least 4 months,
(2) no teacher reported concerns.
(3) no IEP.
All children with LI met the following criteria:
(1) enrolled in speech therapy for at least 4 months prior to participating in this study,
(2) teacher reported concerns.
(3) identified by a bilingual SLP with more than three years of experience as LI.

A parental consent form in Spanish and English was given to all parents. The primary researcher/graduate research assistant explained the research project and all parts of the IRB approved consent forms to the parents. Only children with signed parental consent participated in this study.

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 their learning, cognitive, and social skills.

The student’s portfolio and IEP (for students with LI) was reviewed to assist in the identification of participant groups.

**Participant Ages/Group Means**

The two groups of children were matched by age; analysis of the average age between groups were very similar ($M = 58.53, M = 58.93$).

**Procedures Data Collection and Training**

Data was collected by the Primary Investigator (PI, bilingual speech-language pathologist) and a bilingual speech-language pathology graduate research assistant who was trained to use culturally appropriate assessment measures and consider cultural/linguistic differences, as well as socio-linguistic variables.

The PI provided direct training for the research assistant. A short test was developed and used to measure fidelity, which examined the research assistant’s familiarity and skill with the administration procedures for the various tasks included. The research assistant received a score of 90% on the short answer test. The research assistant also administered portions of all the
subtests in a role-play situation, and the PI observed administration of the test and provided feedback. Besides for administering parts of the assessment measures to the PI in a role-play situation, the research assistant also administered parts of assessment measures in the field before the start of data collection. The research assistant followed through the steps of familiarization, practice, and reliability during the training program. Fidelity was measured and then the indicators re-examined until the graduate assistant achieved at least 90% accuracy on the administration and scoring of all assessment measures. This made certain that the graduate assistant had fidelity to the administration process; she was competent in following the procedures as planned. Throughout the data collection period, the primary researcher also spot-checked to ensure that all of the procedures were continuously administered as planned.

**Evaluation procedures**

After parental consent was obtained, a first year speech-language pathology graduate student and bilingual speech-language pathologist (PI) administered all language assessment measures to all participants in the LI and TL groups. The following assessment measures were given in random order to participants across 4 – 6 sessions, each lasting approximately one hour. The language samples, standardized assessments, mediated learning activities during the DA task, and NWR tasks were also counterbalanced across participants.

All of the assessment measures were administered to the children in a quiet room at their schools.

**Language sample** - A storytelling task was used because it fosters greater complexity than a language sample collected through play (Fiestas & Pena, 2004; Restrepo & Castilla, 2007; Restrepo, et al., 2001). Participants were asked to retell a story that corresponded to the wordless picture book, *Frog Where Are You*, twice after the story was told by the data collector. The
language samples were used as a static language measure and administered in both Spanish and English in order to obtain a Spanish and English language sample from each participant. The participants’ better language sample was included in the data analysis in order to obtain the best estimate of the child’s language abilities. The language samples were transcribed and then MLU-w scores were calculated.

**Standardized assessments** – The following standardized assessments were administered according to the directions stated in the manual: PLS-4 Spanish Edition, PLS-4 English Edition, CELF-P2 Spanish Edition, CELF-P2 English Edition, BESA (semantics and morphology subtests in Spanish and English). Standard scores were calculated for all of the standardized assessments except the BESA because BESA standard scores were not available at the time of data analysis and the test is still in development stages. The data analysis included total language measures for all the standardized tests except that the semantics and morpho-syntax subtests for the BESA were also input for analysis, given that it is a new test and analysis of each subtest in the model could be informative for researchers.

**Parental report** – The parent interview developed by Restrepo (1998) was used because its accuracy to distinguish predominantly Spanish-speaking children with language impairment and typical language learners is close to the minimum set by Plante and Vance (1994). Specifically, parental report of the child’s speech and language problems had a sensitivity value of 73.91% and a specificity value of 95.65%. The sensitivity and specificity values of the parent report of the child’s speech and language problems along with the number of errors per T-unit were greater than 90%, which is above the minimum value of 80% established by Plante and Vance (1994). This parental report collects information based on parents’ interpretation of their child’s overall speech-language skills, as well as information about family history of speech and
language problems. Parents are asked 50 questions (21 about family history, 29 questions about the child’s speech-language skills) and a higher the score indicates that the child presents with more at-risk behaviors and is more likely to have a language impairment. Total raw scores were used for the data analysis, as was done in the original research study (Restrepo, 1998). The parent interview was administered in Spanish since all of the participants’ parents speak primarily Spanish.

Non Word Repetition (NWR) – NWR tasks in Spanish and English were administered. The English and Spanish NWR tasks used by Windsor, et al. (2010) were administered. Both NWR tasks follow the phonotactic constraints of the Spanish and English language respectively. The English NWR task was developed by Dollaghan and Campbell (1998) and consisted of 16 nonwords, four at each length from one to four syllables. The Spanish NWR task was developed by Ebert, Kalanek, Cordero, and Kohnert (2008) and consists of 20 nonwords, with four words for each syllable length of 1-5 syllables.

The nonwords in each task were spoken by native speakers and recorded onto a CD. The primary researcher obtained the recorded nonwords from the original researchers who developed these tasks so that the participants listened to the same recorded nonwords as in previous studies (Dollaghan & Campbell, 1998; Ebert, et al., 2008; Windsor, et al., 2010) The two NWR tasks were administered in counterbalanced order.

Instructions were provided in Spanish for the NWR Spanish task and English for the English NWR task. The two tasks were administered to the children in a quiet room at their schools. The nonwords in both tasks were presented through headphones to the children; however, some children refused to wear the headphones and so they listened to the recording of the nonwords in the quiet room without wearing the headphones. The children were then
expected to provide one imitation of each nonword. Children’s productions were audio-recorded, and each task was scored for percentage of consonants produced correctly (PCC).

**Dynamic assessment (DA)** – A test, teach, retest dynamic assessment procedure as described in Pena, et al. (2001) was used, targeting word learning (vocabulary). In the vocabulary DA task, children were given the EOWPVT Bilingual Spanish Edition as a pretest, participated in two 20 minute mediation sessions, and then completed the EOWPVT Bilingual Spanish Edition as a posttest. During the mediated sessions, four activities included in the Mediated Learning Script (MLE) were administered, which were counterbalanced across participants. Each of these activities focused on teaching children about the principles of labeling through discussion, had children problem solve and plan through activities to promote self-directed learning, and used toy sets centered around themes (foods, transportation, animals and community workers). The data collector stated the goal and purpose of each session, introduced the activity and how it relates to children’s school/home activities, helped children plan and carry through the activity, and provided a closing that summarized the labeling principles and student’s behavior during the activity. Although all of the MLE activities contained the same components, the data collectors were encouraged to be responsive to each student in order to provide individualized feedback and assist the child in learning to label objects.

Since language proficiency is not examined during a dynamic assessment task; rather, the purpose is to assess the language-learning potential of the child, both Spanish and English language productions/responses were accepted during the dynamic assessment task. Data collectors were also flexible in their language use and mostly used the language of the participant during the dynamic assessment task in order to provide the individualized feedback for the participants, which is one aspect of the MLE approach. Thus, the dynamic assessment
intervention activities were functional and simulated a naturalistic environment. Conceptual scoring was used for those students who responded using both languages. Conceptual scoring considers both Spanish and English responses, but the response in either or both languages is only counted once (Bedore, Pena, Garcia, & Cortez, 2005).

The dynamic assessment tasks were calculated in two different ways; the difference between the pretest and posttest standard scores on the EOWPVT Bilingual Spanish Edition and the difference between the pretest and posttest raw scores on the EOWPVT Bilingual Spanish Edition minus the actual words targeted during the mediation sessions were used.

Data Analysis

Reliability

In order to test reliability, the PI (bilingual speech-language pathologist) scored 10% of all assessment measures. The inter-rater reliability was 94% for point to point agreement on each item across all of the assessments. For the language samples, a bilingual speech-language pathologist (primary investigator) transcribed the language samples using the SALT software and then another bilingual speech-language pathologist transcribed 10% of them to assess reliability of the transcription. Inter-rater reliability was 92%.

A priori power analysis was conducted in order to justify the number of subjects selected for the study. In this power analysis, a small effect size of .25 was estimated. In order to reach significance at 80% power with at least a .25 effect size, approximately 34 subjects were predicted to be necessary. A statistically significant result and a large effect size was found after collecting data from 30 participants and so data collection was then discontinued. Although during the priori power analysis only 7 predictors were predicted, it was important to input both English and Spanish scores separately, making them separate predictors and increasing the
number of predictors to a total of 16. However, it was necessary to reduce the number of variables in order to achieve significance with at least 80% power, given a sample size of 30 participants. Therefore, a regression statistical analysis was initially completed with all of the predictors and then a second regression was conducted including the 6 most predictive variables.

Multiple Linear regression analysis was used even though there is a dichotomous dependent variable (DV) and many independent variables (IV) because Hellevik (2009) explains that a violation of the homoscedacity assumption has little effect on the outcome; the results from the significance tests of both linear and logistic regression are identical even when small samples and skewed distributions are used. Given that using a linear regression is non-
problematic when there is a dichotomous dependent variable, it is beneficial to use a linear regression because it is basically interpreting a probability, which is more easily understood than a logistic regression. All the independent variables included produce continuous numeric output.

The two groups of children were matched by age; analysis of the average age between groups are shown in Table 1 and were very similar (M = 58.53, M = 58.93). Thus, comparison of the dataset between TL and LI groups are appropriate.

Table 1

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>M (SD)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>TL</td>
<td>15</td>
<td>58.53 (6.14)</td>
<td>55.13, 61.93</td>
</tr>
<tr>
<td>LI</td>
<td>15</td>
<td>58.93 (6.01)</td>
<td>55.61, 62.26</td>
</tr>
</tbody>
</table>

Note: CI = confidence interval. Age is in months
CHAPTER IV

RESULTS

Two research questions drove this study. The results for these questions are reported separately. In addition, follow-up analysis examining the difference between performance on English and Spanish standardized assessments were completed, and the results are also reported.

Research Question 1: Which types of assessment [standardized assessment (PLS-4 English, PLS-4 Spanish, CELF – P2 English, CELF-P2 Spanish, Bilingual English Spanish Assessment - BESA); Parental Report (Restrepo (1998) parent interview); language samples (English and Spanish – MLU-w); NWR (Spanish and English); dynamic assessment (word-learning ability/vocabulary)] best distinguish between bilingual Spanish-English speaking children with language impairment and typically developing children?

Descriptive data for all of the assessment measures are provided in Table 2. Means and standard deviations were calculated for each assessment measure for the entire sample, as well as for TL and LI groups. All assessment measures were administered to all participants; however, some children either did not produce any language sample or produced a language sample in only one language. Specifically 9 children (5 TL 4 LI) produced both Spanish and English language samples: 7 children (3 TL 4 LI) produced only Spanish language samples: and 12 children (6 TL 6 LI) produced only English language samples. Two participants did not complete the Spanish NWR task; thus, a Percentage of Consonants Correct (PCC) score was not calculated for them, and this was considered missing data.

A multiple linear regression was used to determine which types of assessment measures best distinguished bilingual Spanish-English speaking children with language impairment from typically developing children. Visual analyses of the point biserial correlations between
assessment measures and language status, as well as statistical results were examined. This research question was examined in two parts although the same regression analyses were used for both; a comparison of standardized assessments was initially examined to determine whether there was a standardized assessment that was better for this population, and then all assessment measures, including authentic assessments, were analyzed to determine which one best distinguished the groups and the predictive value of each of them.

Table 2

Means, Standard Deviations, and Confidence Intervals for Assessment Measures
### Table 1: Measures of Language Proficiency for Bilingual Spanish-English Speaking Children with Language Impairment

<table>
<thead>
<tr>
<th>Measure</th>
<th>TD 15</th>
<th>LI 15</th>
<th>Total 30</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PLS4 English</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 15</td>
<td>92.93</td>
<td>72.27</td>
<td>82.60</td>
</tr>
<tr>
<td>LI 15</td>
<td>10.40</td>
<td>12.26</td>
<td>15.34</td>
</tr>
<tr>
<td>TD 15</td>
<td>3.17</td>
<td>6.47</td>
<td>2.80</td>
</tr>
<tr>
<td>LI 15</td>
<td>87.17</td>
<td>3.68</td>
<td>76.87</td>
</tr>
<tr>
<td>TD 15</td>
<td>98.69</td>
<td>98.69</td>
<td>88.33</td>
</tr>
<tr>
<td>LI 15</td>
<td>80</td>
<td>109</td>
<td>109</td>
</tr>
<tr>
<td><strong>PLS4 Spanish</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 15</td>
<td>103.07</td>
<td>69.80</td>
<td>86.43</td>
</tr>
<tr>
<td>LI 15</td>
<td>12.68</td>
<td>15.35</td>
<td>21.85</td>
</tr>
<tr>
<td>TD 15</td>
<td>3.27</td>
<td>3.96</td>
<td>2.80</td>
</tr>
<tr>
<td>LI 15</td>
<td>65.47</td>
<td>61.31</td>
<td>63.84</td>
</tr>
<tr>
<td>TD 15</td>
<td>110.09</td>
<td>79.06</td>
<td>94.59</td>
</tr>
<tr>
<td>LI 15</td>
<td>53</td>
<td>87</td>
<td>53</td>
</tr>
<tr>
<td><strong>CELF-P2 English</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 15</td>
<td>74.93</td>
<td>69.73</td>
<td>86.43</td>
</tr>
<tr>
<td>LI 15</td>
<td>11.06</td>
<td>10.97</td>
<td>21.85</td>
</tr>
<tr>
<td>TD 15</td>
<td>2.86</td>
<td>2.83</td>
<td>2.80</td>
</tr>
<tr>
<td>LI 15</td>
<td>68.81</td>
<td>63.65</td>
<td>63.84</td>
</tr>
<tr>
<td>TD 15</td>
<td>81.06</td>
<td>75.81</td>
<td>77.01</td>
</tr>
<tr>
<td>LI 15</td>
<td>57</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td><strong>CELF-P2 Spanish</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 15</td>
<td>97.93</td>
<td>69.73</td>
<td>83.83</td>
</tr>
<tr>
<td>LI 15</td>
<td>12.68</td>
<td>10.97</td>
<td>21.85</td>
</tr>
<tr>
<td>TD 15</td>
<td>3.16</td>
<td>2.83</td>
<td>2.80</td>
</tr>
<tr>
<td>LI 15</td>
<td>91.25</td>
<td>75.81</td>
<td>77.01</td>
</tr>
<tr>
<td>TD 15</td>
<td>104.62</td>
<td>75.81</td>
<td>90.66</td>
</tr>
<tr>
<td>LI 15</td>
<td>73</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td><strong>Dynamic Assessment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 15</td>
<td>8.53</td>
<td>3.33</td>
<td>5.93</td>
</tr>
<tr>
<td>LI 15</td>
<td>12.90</td>
<td>10.97</td>
<td>21.85</td>
</tr>
<tr>
<td>TD 15</td>
<td>3.33</td>
<td>2.83</td>
<td>2.80</td>
</tr>
<tr>
<td>LI 15</td>
<td>63.65</td>
<td>63.65</td>
<td>63.84</td>
</tr>
<tr>
<td>TD 15</td>
<td>104.62</td>
<td>75.81</td>
<td>90.66</td>
</tr>
<tr>
<td>LI 15</td>
<td>73</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td><strong>EOWPVTRSS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 15</td>
<td>2.27</td>
<td>1.60</td>
<td>5.53</td>
</tr>
<tr>
<td>LI 15</td>
<td>5.53</td>
<td>4.553</td>
<td>.831</td>
</tr>
<tr>
<td>TD 15</td>
<td>1.43</td>
<td>.831</td>
<td>.68</td>
</tr>
<tr>
<td>LI 15</td>
<td>.70</td>
<td>.83</td>
<td>.68</td>
</tr>
<tr>
<td>TD 15</td>
<td>5.33</td>
<td>3.30</td>
<td>5.53</td>
</tr>
<tr>
<td>LI 15</td>
<td>.80</td>
<td>.80</td>
<td>.80</td>
</tr>
<tr>
<td><strong>NWR PCC English</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 15</td>
<td>.76</td>
<td>.67</td>
<td>.103</td>
</tr>
<tr>
<td>LI 15</td>
<td>.133</td>
<td>.176</td>
<td>.034</td>
</tr>
<tr>
<td>TD 15</td>
<td>.68</td>
<td>.61</td>
<td>.034</td>
</tr>
<tr>
<td>LI 15</td>
<td>.68</td>
<td>.74</td>
<td>.68</td>
</tr>
<tr>
<td>TD 15</td>
<td>.83</td>
<td>.74</td>
<td>.68</td>
</tr>
<tr>
<td>LI 15</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>NWR PCC Spanish</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 14</td>
<td>.83</td>
<td>.75</td>
<td>.103</td>
</tr>
<tr>
<td>LI 14</td>
<td>.103</td>
<td>.136</td>
<td>.027</td>
</tr>
<tr>
<td>TD 15</td>
<td>.77</td>
<td>.70</td>
<td>.027</td>
</tr>
<tr>
<td>LI 15</td>
<td>.70</td>
<td>.80</td>
<td>.027</td>
</tr>
<tr>
<td>TD 15</td>
<td>.89</td>
<td>.80</td>
<td>.027</td>
</tr>
<tr>
<td>LI 15</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Lang. Sample</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 14</td>
<td>6.20</td>
<td>.75</td>
<td>1.76</td>
</tr>
<tr>
<td>LI 14</td>
<td>1.76</td>
<td>.136</td>
<td>.456</td>
</tr>
<tr>
<td>TD 15</td>
<td>5.21</td>
<td>.70</td>
<td>.456</td>
</tr>
<tr>
<td>LI 15</td>
<td>7.18</td>
<td>.80</td>
<td>.456</td>
</tr>
<tr>
<td>TD 15</td>
<td>4.28</td>
<td>1</td>
<td>4.28</td>
</tr>
<tr>
<td>LI 15</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td><strong>MLU-w</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TD 14</td>
<td>4.84</td>
<td>28</td>
<td>2.053</td>
</tr>
<tr>
<td>LI 14</td>
<td>1.374</td>
<td>28</td>
<td>.388</td>
</tr>
<tr>
<td>TD 15</td>
<td>4.04</td>
<td>28</td>
<td>4.04</td>
</tr>
<tr>
<td>LI 15</td>
<td>5.64</td>
<td>28</td>
<td>5.64</td>
</tr>
<tr>
<td>TD 15</td>
<td>1</td>
<td>28</td>
<td>1</td>
</tr>
<tr>
<td>LI 15</td>
<td>7</td>
<td>28</td>
<td>7</td>
</tr>
</tbody>
</table>

**Note:** On the Restrepo (1998) parent interview, the lower parent interview score indicates that participants had fewer at-risk behaviors, and thus is then less likely to have a language impairment.

Multicollinearity diagnostics were assessed to determine if the predictor variables were interrelated. Field (2005) states that:

“One way of identifying multicollinearity is to scan a correlation matrix of all the predictor variables and see if any correlate very highly (correlations above .80 or .90).”
Correlations between assessment measures are summarized in Table 6. Most of the assessment measures’ correlations were interrelated, but they were less than .90. SPSS collinearity statistics were also performed and the results show that there is some collinearity between most of the variables; however, this is not surprising as they all are testing the same factors. Although most of the predictors were significantly correlated with Language Status, they met tolerance for multicollinearity and could be examined separately.

Multiple linear regression analysis was performed on the dataset to determine which predictor variables were significant indicators of a language impairment, as measured by the criterion variables. Table 3 presents the correlation matrix for the regression of assessment measures predicting language status. Only the BESA Spanish, CELF-P2 Spanish, and PLS-4 Spanish had strong correlations with language status. The Restrepo (1998) parent interview, BESA English, CELF-P2 English, PLS-4 English, NWR Spanish, and Language Sample tasks had moderate correlations and the NWR English had a weak correlation with language status. Note that the negative correlations for statistically significant correlations indicate identification of a language impairment on all assessment measures except for the parent interview which shows a positive correlation for identification of a language impairment. The correlation for language status/group membership and all assessment measures except for the parent interview are negative. Thus, as children’s’ performance on any of these assessment measures decreases, they are more likely to be identified as having a language impairment. The higher parent interview scores indicate that the child has more speech/language difficulties and is more likely to have a language impairment.
Point Biserial Correlations for Participants’ (N=30) Between Assessment Measures and Language Status

<table>
<thead>
<tr>
<th>Language status</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Restrepo Total</td>
<td>.577</td>
</tr>
<tr>
<td>BESA English Semantics</td>
<td>-.503</td>
</tr>
<tr>
<td>BESA English Syntax</td>
<td>-.563</td>
</tr>
<tr>
<td>BESA English Total</td>
<td>-.561</td>
</tr>
<tr>
<td>BESA Spanish Semantics</td>
<td>-.587</td>
</tr>
<tr>
<td>BESA Spanish Syntax</td>
<td>-.808</td>
</tr>
<tr>
<td>BESA Spanish Total</td>
<td>-.780</td>
</tr>
<tr>
<td>PLS-4 English</td>
<td>-.685</td>
</tr>
<tr>
<td>PLS-4 Spanish</td>
<td>-.774</td>
</tr>
<tr>
<td>CELF-P2 English</td>
<td>-.586</td>
</tr>
<tr>
<td>CELF-P2 Spanish</td>
<td>-.785</td>
</tr>
<tr>
<td>Dynamic Assessment</td>
<td>-.230</td>
</tr>
<tr>
<td>EOWPVT SS</td>
<td></td>
</tr>
<tr>
<td>Dynamic Assessment</td>
<td>-.149</td>
</tr>
<tr>
<td>EOWPVT RS</td>
<td></td>
</tr>
<tr>
<td>NWR English</td>
<td>-.493</td>
</tr>
<tr>
<td>NWR Spanish</td>
<td>-.643</td>
</tr>
<tr>
<td>LS MLU-w</td>
<td>-.672</td>
</tr>
</tbody>
</table>

All variables were entered into the equation stepwise. The overall regression was F(15,10) = 5.762, p < .05. Table 4 presents the results of the multiple regression analysis. Although all variables except dynamic assessment were statistically significant, the BESA Spanish Morphosyntax subtest was the most highly discriminating in predicting language impairment: beta = -.808, t(10) = -3.879, p < .01. Almost 90% of the variance in the model is explained by the predictors of the parent interview, English and Spanish standardized assessments, English and Spanish NWR, and language sample (R² = .896, F (17, 8) = 4.097, p < .05). All measures except the dynamic assessment task can be used to assist in identification of children with language impairment; however, the BESA Spanish assessment may be the most discriminating for this population because it explained most of the variability. To answer part 1 of the first research question, all Spanish standardized assessments were highly discriminating
and showed strong correlations with language status. Although the English standardized assessments were significant, they had only moderate correlations with language status. To answer part 2 of the first research question, the Spanish standardized assessments were highly discriminating and had strong correlations with language status while the parent interview, English standardized assessments, NWR Spanish and language sample were highly discriminating and had moderate correlations with language status. Regression analysis was calculated with MLU-w (mean length of utterance for words) and MLU-m (mean length of utterance for morphemes) separately and both showed moderate correlations with language status. Since MLU-w (-.672) had a slightly stronger correlation with language status than MLU-m (-.656), the MLU-w score was the measurement used for the language sample and input as a predictor in the regression. Although the NWR English task was statistically significant, it had only a weak correlation with language status.

Table 4

*Multiple Linear Regression Analyses for Predictors of Language Status*

<table>
<thead>
<tr>
<th></th>
<th>Model 1 B</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.95*</td>
<td>[.02, 5.88]</td>
</tr>
<tr>
<td>Restrepo Total</td>
<td>.01</td>
<td>[-.02, .03]</td>
</tr>
<tr>
<td>BESA English Semantics</td>
<td>-.09</td>
<td>[-.21, .04]</td>
</tr>
<tr>
<td>BESA English Syntax</td>
<td>-.05</td>
<td>[-.14, .04]</td>
</tr>
<tr>
<td>BESA English Total</td>
<td>.06</td>
<td>[-.04, .16]</td>
</tr>
<tr>
<td>BESA Spanish Semantics</td>
<td>.07</td>
<td>[-.04, .17]</td>
</tr>
<tr>
<td>BESA Spanish Total</td>
<td>-.02</td>
<td>[-.06, .02]</td>
</tr>
</tbody>
</table>


In order to investigate whether there was statistical significance with at least 80% power, the multiple linear regression analysis was repeated with only the six highest predictors included. All variables were entered into the equation stepwise. The overall regression was $F(6, 19) = 7.132, p < .001$. Table 5 presents the results of the multiple regression analysis. All the variables were statistically significant. The CELF-P2 Spanish was the most highly discriminating in predicting language impairment, but the BESA Spanish Total and PLS-4 Spanish were correlated with language status at about the same values. Almost 69% of the variance in the model is explained by the BESA Spanish Total, PLS-4 English, PLS-4 Spanish, CELF-P2 Spanish, NWR Spanish and Language Sample ($R^2 = .693, F(6, 19) = 7.132, p < .01$). Since the results from the regression analysis with fewer predictors were similar to the first regression, including all of the predictors, there is confidence that the initial regression is also appropriate, and the interpretations are accurate.
**Table 5**

*Multiple Linear Regression Analyses for the 6 Highest Predictors of Language Status*

<table>
<thead>
<tr>
<th></th>
<th>Model 1 B</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>3.37**</td>
<td>[2.13, 4.61]</td>
</tr>
<tr>
<td>BESA Spanish Total</td>
<td>-.01</td>
<td>[-.02, .01]</td>
</tr>
<tr>
<td>PLS-4 English</td>
<td>-.01</td>
<td>[-.02, .01]</td>
</tr>
<tr>
<td>PLS-4 Spanish</td>
<td>-.01</td>
<td>[-.03, .01]</td>
</tr>
<tr>
<td>CELF-P2 Spanish</td>
<td>-.00</td>
<td>[-.03, .02]</td>
</tr>
<tr>
<td>NWR Spanish</td>
<td>-.09</td>
<td>[-1.92, 1.75]</td>
</tr>
<tr>
<td>LS MLU-w</td>
<td>-.07</td>
<td>[-.19, .05]</td>
</tr>
</tbody>
</table>

*Note. N = 30. CI = confidence interval.*

**p < .01

Research Question 2: Is there a group of tests that best distinguishes between bilingual Spanish-English speaking children with language impairment and typically developing children [standardized assessment (PLS-4 English, PLS-4 Spanish, CELF – P2 English, CELF-P2 Spanish, Bilingual English Spanish Assessment - BESA); Parental Report (Restrepo (1998) parent interview); language samples (English and Spanish – MLU-w); NWR (Spanish and English); dynamic assessment (word-learning ability/vocabulary)]?*

Further examination of the data was conducted to examine the correlations between assessment measures (not only the correlations between the dependent and independent variables) in order to determine which assessment measures as a group best distinguish between bilingual Spanish-English speaking children with language impairment and those with typical
language. Point biserial correlations between assessment measures were reviewed to determine which assessments do or do not have strong correlations with each other, as well as whether those tests also highly discriminated TL from LI groups. If there are assessment measures that highly discriminate TL from LI groups and do not correlate strongly with other assessment measures, these assessments may be grouped together and used as a comprehensive battery to identify a language impairment. On the other hand, tests that are highly discriminating but strongly correlated with each other seem to measure the same thing, and only one should be included in a battery. A correlation greater than .7 was considered to be strong, .5 to .6 was considered moderate, and .4 was considered a weak correlation. When using a regression statistical analysis, a statistically significant result might be found; however, it is important to also examine the correlations to determine if a practical effect actually exists.

Table 6 presents the correlation matrix of all variables included in the data. The BESA Spanish, CELF-P2 Spanish, and PLS-4 Spanish were strongly correlated and highly discriminating, suggesting that these assessments measure similar linguistic skills, and therefore administration of only one of these would be sufficient. The English standardized assessments (BESA English, CELF-P2 English, and PLS-4 English) were also strongly correlated with each other and highly discriminating; thus, only one of these assessments would be sufficient. [Each standardized assessment was also strongly correlated with its counterpart (e.g., CELF-P2 English and CELF-P2 Spanish).] The Parent interview is not strongly correlated with the other predictors, but is highly discriminating suggesting that it assesses different areas and should be included in a battery. The Language Sample was strongly correlated with the BESA Spanish and CELF-P2 Spanish. The Spanish NWR was strongly correlated with the BESA Spanish and CELF-P2 Spanish, and PLS-4 Spanish ($r = .69, p < .01$). Only a weak correlation was found
between the English NWR task and language status, suggesting that a practical/meaningful effect does not exist even though the English NWR task was statistically significant. The dynamic assessment task was not statistically significant and did not correlate strongly with any of the other predictors.

Since all of the predictors except the dynamic assessment task were highly discriminating, group assignment and assessment performance should correspond, meaning that the participants are assumed to have been assigned to the appropriate group within the study. In order to confirm this information, sensitivity and specificity values were calculated for the battery suggested. The criteria used were that the participant’s scores for at least two of three assessment measures matched the group as assignment. For example, a child in the TL group who performed within 1 SD of the mean on the PLS-4 Spanish and above the NWR PCC cutoff score (80%) was considered to be accurately identified. The discriminant model classified 13 of 15 of the children in the LI group as LI, indicating that the model had a sensitivity of 86.6%; whereas it classified 14 of 15 children in the TL group as TL, indicating a specificity of 93.3%. These values are above the minimum established by Plante and Vance (1994) indicating that the battery suggested in this study is appropriate to identify a LI in 4-5 year old bilingual Spanish-English speaking children accurately.
### Table 6

**Point Biserial Correlations Between Assessment Measures**

<table>
<thead>
<tr>
<th></th>
<th>Restrepo parent interview</th>
<th>BESA Eng</th>
<th>BESA Eng</th>
<th>BESA Eng</th>
<th>BESA Sp</th>
<th>BESA Sp</th>
<th>BESA Sp</th>
<th>PLS4 Eng</th>
<th>PLS4 Sp</th>
<th>CELF-P2 Eng</th>
<th>CELF-P2 BSE-SS</th>
<th>CELF-P2 BSE-RS</th>
<th>EOWPVT</th>
<th>EOWPVT BSE-SS</th>
<th>EOWPVT BSE-RS</th>
<th>NWR PCC Eng</th>
<th>NWR PCC Sp</th>
<th>LS MLUw</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Restrepo parent interview</td>
<td>-.58**</td>
<td>-.50**</td>
<td>-.56**</td>
<td>-.56**</td>
<td>-.59**</td>
<td>-.81**</td>
<td>-.78**</td>
<td>-.69**</td>
<td>-.77**</td>
<td>-.59**</td>
<td>-.79**</td>
<td>-.23</td>
<td>-.15</td>
<td>-.49**</td>
<td>-.64**</td>
<td>-.67**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BESA Eng</td>
<td>-25</td>
<td>-.76**</td>
<td>.87**</td>
<td>.76**</td>
<td>.56**</td>
<td>.65**</td>
<td>.77**</td>
<td>.58**</td>
<td>.79**</td>
<td>.45*</td>
<td>.19</td>
<td>-.28</td>
<td>.60**</td>
<td>.47*</td>
<td>.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BESA Eng Syntax</td>
<td>-22**</td>
<td>.76**</td>
<td>--</td>
<td>.98**</td>
<td>.74**</td>
<td>.68**</td>
<td>.73**</td>
<td>.73**</td>
<td>.63**</td>
<td>.85**</td>
<td>.58**</td>
<td>.34</td>
<td>.38</td>
<td>.66</td>
<td>.56**</td>
<td>.47*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BESA Sp</td>
<td>-24**</td>
<td>.87**</td>
<td>.98**</td>
<td>--</td>
<td>.79**</td>
<td>.68**</td>
<td>.75**</td>
<td>.78**</td>
<td>.64**</td>
<td>.87**</td>
<td>.57**</td>
<td>.31</td>
<td>.37</td>
<td>.66</td>
<td>.55**</td>
<td>.47*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BESA Sp Syntax</td>
<td>-.41*</td>
<td>.76**</td>
<td>.74**</td>
<td>.79**</td>
<td>--</td>
<td>.78**</td>
<td>.89**</td>
<td>.72**</td>
<td>.74**</td>
<td>.63**</td>
<td>.68**</td>
<td>.10</td>
<td>.13</td>
<td>.67**</td>
<td>.68**</td>
<td>.65**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BESA Sp Semantics</td>
<td>-.64**</td>
<td>.56**</td>
<td>.68**</td>
<td>.68**</td>
<td>.78**</td>
<td>--</td>
<td>.98**</td>
<td>.61**</td>
<td>.83**</td>
<td>.58**</td>
<td>.87**</td>
<td>.08</td>
<td>.09</td>
<td>.59**</td>
<td>.72**</td>
<td>.67**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BESA Sp Total</td>
<td>-.60**</td>
<td>.65**</td>
<td>.73**</td>
<td>.75**</td>
<td>.89**</td>
<td>.98**</td>
<td>--</td>
<td>.68**</td>
<td>.84**</td>
<td>.62**</td>
<td>.86**</td>
<td>.09</td>
<td>.10</td>
<td>.64**</td>
<td>.74**</td>
<td>.70**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS4 Eng</td>
<td>-.30**</td>
<td>.77**</td>
<td>.73**</td>
<td>.72**</td>
<td>.72**</td>
<td>.61**</td>
<td>.68**</td>
<td>--</td>
<td>.74**</td>
<td>.82**</td>
<td>.60**</td>
<td>.41</td>
<td>.44*</td>
<td>.54**</td>
<td>.52**</td>
<td>.55**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS4 Sp</td>
<td>-.48**</td>
<td>.58**</td>
<td>.63**</td>
<td>.64**</td>
<td>.74**</td>
<td>.83**</td>
<td>.84**</td>
<td>.74**</td>
<td>--</td>
<td>.67**</td>
<td>.88**</td>
<td>.18</td>
<td>.19</td>
<td>.60**</td>
<td>.69**</td>
<td>.52**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CELF-P2 Eng</td>
<td>-.14**</td>
<td>.80**</td>
<td>.85**</td>
<td>.87**</td>
<td>.63**</td>
<td>.58**</td>
<td>.62**</td>
<td>.82**</td>
<td>.67**</td>
<td>--</td>
<td>.59**</td>
<td>.40</td>
<td>.49**</td>
<td>.63**</td>
<td>.56**</td>
<td>.39*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CELF-P2 Sp</td>
<td>-.53**</td>
<td>.45*</td>
<td>.58**</td>
<td>.57**</td>
<td>.68**</td>
<td>.87**</td>
<td>.86**</td>
<td>.60**</td>
<td>.88**</td>
<td>.59**</td>
<td>--</td>
<td>.08</td>
<td>.08</td>
<td>.61**</td>
<td>.77**</td>
<td>.72**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOWPVT BSE-SS</td>
<td>.03</td>
<td>.19</td>
<td>.34</td>
<td>.31</td>
<td>.10</td>
<td>.08</td>
<td>.09</td>
<td>.41*</td>
<td>.18</td>
<td>.40*</td>
<td>.08</td>
<td>--</td>
<td>.85**</td>
<td>-.05</td>
<td>-.14</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOWPVT BSE-RS</td>
<td>.04</td>
<td>.23</td>
<td>.38*</td>
<td>.37*</td>
<td>.13</td>
<td>.09</td>
<td>.10</td>
<td>.44*</td>
<td>.19</td>
<td>.49**</td>
<td>.08</td>
<td>.85**</td>
<td>--</td>
<td>.04</td>
<td>.09</td>
<td>-.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NWR PCC Eng</td>
<td>-.06**</td>
<td>.60**</td>
<td>.66**</td>
<td>.66**</td>
<td>.67**</td>
<td>.59**</td>
<td>.64**</td>
<td>.54**</td>
<td>.60**</td>
<td>.63**</td>
<td>.61**</td>
<td>-.05</td>
<td>.04</td>
<td>--</td>
<td>.81**</td>
<td>.50**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NWR PCC Sp</td>
<td>-.35**</td>
<td>.47*</td>
<td>.56**</td>
<td>.55**</td>
<td>.66**</td>
<td>.72**</td>
<td>.74**</td>
<td>.52**</td>
<td>.69**</td>
<td>.56**</td>
<td>.77**</td>
<td>-.14</td>
<td>-.09</td>
<td>.81**</td>
<td>.59**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS MLUw</td>
<td>-.39**</td>
<td>.37</td>
<td>.47*</td>
<td>.50*</td>
<td>.65**</td>
<td>.67**</td>
<td>.70**</td>
<td>.55**</td>
<td>.52**</td>
<td>.39*</td>
<td>.72**</td>
<td>-.04</td>
<td>-.04</td>
<td>.50**</td>
<td>.59**</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Additional Analysis

A statistical MANOVA was conducted to investigate whether there was a difference in performance on the English and Spanish standardized assessment measures. Although this was not a research question at the outset of this study, a follow-up analysis was appropriate after finding that although both Spanish and English standardized assessments were highly discriminating, the Spanish versions were strongly correlated and the English versions were only moderately correlated with language status. Table 7 shows that there were significant main effects for Language Status, $F_A(1, 60) = 24.044$, $p<.001$ and for Test Language, $F_B(1, 60) = 24.881$, $p<.0001$. Table 8 indicates a significant Language Status by Test Language interaction for the CELF-P2: $F_{AB}(1, 60) = 7.212$, $p<.05$. Specifically, participants performed higher on the CELF-P2 Spanish than the CELF-P2 English.

Table 7

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pillai's Trace</th>
<th>Wilks' Lambda</th>
<th>Hotelling's Trace</th>
<th>Roy's Largest Root</th>
<th>Hypothesis df</th>
<th>Error df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.988</td>
<td>1457.330</td>
<td>3.000</td>
<td>54.000</td>
<td></td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.012</td>
<td>1457.330a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Langstatus</td>
<td>.572</td>
<td>24.044a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.428</td>
<td>24.044a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.336</td>
<td>24.044a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.336</td>
<td>24.044a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>testlang</td>
<td>.580</td>
<td>24.881a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.420</td>
<td>24.881a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.382</td>
<td>24.881a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.382</td>
<td>24.881a</td>
<td>3.000</td>
<td>54.000</td>
<td>.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Langstatus * testlang</td>
<td>.144</td>
<td>3.038a</td>
<td>3.000</td>
<td>54.000</td>
<td>.037</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Multivariate Analysis of Variance Effects of Test Language and Language Status Group
Table 8

Multivariate Analysis of Variance of Assessment Performance by Test Language and Language Status Group

<table>
<thead>
<tr>
<th>Source</th>
<th>Dependent Variable</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected Model</td>
<td>BESAEngTot</td>
<td>10001.650&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3</td>
<td>3333.883</td>
<td>15.367</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>PLS4Eng</td>
<td>11723.783&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3</td>
<td>3907.928</td>
<td>23.872</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>CELFP2Eng</td>
<td>10990.450&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3</td>
<td>3663.483</td>
<td>32.459</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>BESAEngTot</td>
<td>71070.417</td>
<td>1</td>
<td>71070.417</td>
<td>327.596</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>PLS4Eng</td>
<td>428584.017</td>
<td>1</td>
<td>428584.017</td>
<td>2618.106</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>CELFP2Eng</td>
<td>346712.017</td>
<td>1</td>
<td>346712.017</td>
<td>3071.873</td>
<td>.000</td>
</tr>
<tr>
<td>Langstatus</td>
<td>BESAEngTot</td>
<td>9804.817</td>
<td>1</td>
<td>9804.817</td>
<td>45.195</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>PLS4Eng</td>
<td>10908.017</td>
<td>1</td>
<td>10908.017</td>
<td>66.634</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>CELFP2Eng</td>
<td>6510.417</td>
<td>1</td>
<td>6510.417</td>
<td>57.682</td>
<td>.000</td>
</tr>
<tr>
<td>testlang</td>
<td>BESAEngTot</td>
<td>40.017</td>
<td>1</td>
<td>40.017</td>
<td>.184</td>
<td>.669</td>
</tr>
<tr>
<td></td>
<td>PLS4Eng</td>
<td>220.417</td>
<td>1</td>
<td>220.417</td>
<td>1.346</td>
<td>.251</td>
</tr>
<tr>
<td></td>
<td>CELFP2Eng</td>
<td>3666.017</td>
<td>1</td>
<td>3666.017</td>
<td>32.481</td>
<td>.000</td>
</tr>
<tr>
<td>Langstatus * testlang</td>
<td>BESAEngTot</td>
<td>156.817</td>
<td>1</td>
<td>156.817</td>
<td>.723</td>
<td>.399</td>
</tr>
<tr>
<td></td>
<td>PLS4Eng</td>
<td>595.350</td>
<td>1</td>
<td>595.350</td>
<td>3.637</td>
<td>.062</td>
</tr>
<tr>
<td></td>
<td>CELFP2Eng</td>
<td>814.017</td>
<td>1</td>
<td>814.017</td>
<td>7.212</td>
<td>.010</td>
</tr>
<tr>
<td>Error</td>
<td>BESAEngTot</td>
<td>12148.933</td>
<td>56</td>
<td>216.945</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PLS4Eng</td>
<td>9167.200</td>
<td>56</td>
<td>163.700</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CELFP2Eng</td>
<td>6320.533</td>
<td>56</td>
<td>112.867</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>BESAEngTot</td>
<td>93221.000</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PLS4Eng</td>
<td>449475.000</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CELFP2Eng</td>
<td>364023.000</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>BESAEngTot</td>
<td>22150.583</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>df</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-----------</td>
<td>----</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLS4Eng</td>
<td>20890.983</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CELFP2Eng</td>
<td>17310.983</td>
<td>59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. $R^2 = .452$ (Adjusted $R^2 = .422$)
b. $R^2 = .561$ (Adjusted $R^2 = .538$)
c. $R^2 = .635$ (Adjusted $R^2 = .615$)
CHAPTER V
DISCUSSION

Speech-language assessment of bilingual Spanish-English speaking children in the United States has been difficult for researchers and clinicians because of the lack of an optimal assessment measure that can accurately differentiate children with typical language and those who are language impaired. In addition, assessment is complicated by typical bilingual language learning patterns children experience when exposed to L2 (second language) that appear similar to the characteristics of a child with a language impairment. It is a challenge to tease out whether the language challenges a child displays are due to the learning of two languages and/or language difficulties. The purpose driving this study was to determine the most appropriate assessment measures to best differentiate bilingual typical language learners and children with language impairment, using assessment measures available and recommended for use with bilingual Spanish-English speaking children. Results from this study can provide information regarding the assessment measures that can be used as a battery for speech-language assessment for this population and provide information for the development of a test that incorporates these types of assessment tasks. The two research questions explored (1) whether a specific assessment measure is best in differentiating typical language learners and children with language impairment; and (2) whether there are a group of assessment measures that collectively best differentiate between the TL (typical language) and LI (language impairment) groups of children. To answer these research questions, five assessment tasks that have been shown to be useful for this population were administered to 4-5 year old children who were typical language learners and those who were language impaired. These included a parent interview, dynamic assessment task, NWR task, language sample, and three standardized assessments. Both English
and Spanish versions of these assessment measures (except for dynamic assessment) were administered because of the varied levels of proficiency that exist among bilingual Spanish-English speaking children, making it necessary to assess these children in both languages to determine the presence/absence of a language impairment. The scores for each assessment measure were then used to determine which assessment measures differentiated between the TL and LI groups. Standard scores were used for the standardized assessments except for the BESA, which is in the development stage, and thus only raw scores were available. PCC (percentage of consonant correct), MLU-w/MLU-m, and pretest/posttest difference scores were used for other assessment measures.

Assessments for the Identification of LI in Bilingual Children

The primary aim of this study was to determine the best assessment measures to be used in the identification of young bilingual children with language impairment. Since a regression analysis is based on correlations, both a statistically significant result, as well as a strong correlation between the assessment measure and group membership should exist to determine the most appropriate assessments for practical purposes. In addition to determining the most effective measures, it is also important to consider efficiency when identifying a child with a language impairment. It would be helpful to determine the fewest assessment measures with the highest level of confidence needed to differentiate TL and LI children. Although all of the predictors except dynamic assessment were highly discriminating, only the BESA Spanish, CELF-P2 Spanish, and PLS-4 Spanish had strong correlations with language status. Therefore it would be recommended that a clinician use the BESA Spanish, CELF-P2 Spanish, or PLS-4 Spanish when assessing young bilingual Spanish-English speaking children from a similar population as in this study (whose native language is Spanish and began formal instruction of
English at the start of school) because it provides the best information to classify a TL or LI child in comparison to the other assessment measures.

The Restrepo (1998) parent interview, BESA English, CELF-P2 English, PLS-4 English, NWR Spanish task and Language Sample had moderate correlations with group membership. These assessments can be used as part of a test battery, but the smaller correlations suggest that the level of confidence would not be as high, and the decision of TL or LI may not be as reliable with only one of these measures. They may be useful as a supplement to provide additional useful information during a speech-language evaluation. The parent interview provides valuable information about the child’s family history of speech-language difficulties and risk factors that might contribute to the presence of a language impairment. Although the Spanish versions of the standardized tests were deemed to be most effective, administration of an English standardized assessment would also be beneficial when it seems important to assess a bilingual child in both languages to determine the level of proficiency in English and Spanish.

Both NWR tasks were statistically significant, but the NWR Spanish had a moderate correlation and the NWR English had a weak correlation with language status. Since NWR is only a co-clinical marker with SLI and strongly correlated with the Spanish standardized assessments, it seems that there is no additional information to assist in the identification of language impairment collected from the NWR Spanish for this population. There is also no practical purpose in administering the NWR English task as a measure of language for this population because it had a weak correlation with language status. However, NWR tasks (English and Spanish) may be useful for describing a child’s articulation skills, examining speech sound production. It should be noted that some of the children included in this study who had a primary LI also had some articulation difficulties.
The Language Sample provides additional qualitative information that cannot be obtained from the other quantitative variables. In comparison to standardized assessments assessing isolated skills in a decontextualized format, language sampling provides information about the child’s spontaneous language skills in naturalistic contexts so that a clinician can use clinical judgment to assist in the identification of a language impairment.

The Dynamic Assessment task did not show statistical significance in discriminating between the two groups. In a response to intervention (Rti) model, a child’s learning potential is at the core of deciding whether a child has a disability; a child is only diagnosed with a language-learning disability if he/she fails to learn when high quality language instruction was provided. The clinician attempts to determine how much assistance the child requires to acquire language skill/s targeted. Assessment measures are also used for intervention planning. Although the results from this study suggest that the dynamic assessment task for word learning cannot be used for the identification of a language impairment, it is an important assessment measure for planning intervention once a child has been diagnosed with a language impairment. It would be valuable information for a clinician to complete a dynamic assessment task to examine the child’s learning potential, level of prompting, and strategies needed to assist a child in learning language skills targeted during intervention.

In this study, only the Spanish standardized assessments met the requirements of statistical significance and were strongly correlated with language status. This was an expected finding as the children in this study had more Spanish than English exposure. Although some of the participants had a higher MLU-w on the English language samples, it seems that overall the children still comprehended/used more linguistic skills in Spanish. This will be discussed further.
in a later section of the discussion (Participants’ Age and Sociolinguistic Factors in Relationship to Assessment Performance).

**Appropriate Battery for the Identification of Language Impairment in Bilingual Children**

A second aim of this study was to determine which group of assessment measures can be combined as a comprehensive battery for the identification of language impairment among young bilingual Spanish-English speaking children. If there are assessment measures that are highly discriminating TL from LI groups and do not correlate strongly with other assessment measures, it is recommended that these assessments be grouped and used as a comprehensive battery to identify a language impairment.

As mentioned earlier, it is reasonable to assume that assessment measures that correlate strongly with each other are assessing the same language area and that assessment measures that do not correlate with each other assess different areas. The BESA Spanish, CELF-P2 Spanish, and PLS-4 Spanish, were strongly correlated and highly discriminating, suggesting that these assessments measure similar linguistic skills. The BESA English, CELF-P2 English, and PLS-4 English, were also strongly correlated and highly discriminating. Therefore, administration of only one of these assessments in each language would be sufficient in a battery. The Parent interview is not strongly correlated with other assessments, but is highly discriminating suggesting that it assesses different areas and should be included in a battery. The Language Sample was strongly correlated with the BESA Spanish and CELF-P2 Spanish. These results suggest that it is only necessary to include a language sample in a battery when the PLS-4 Spanish is administered. However, a language sample would supplement any standardized assessment (BESA Spanish, CELF-P2 Spanish) with qualitative information to assist in identifying and planning intervention for a child with a language impairment (discussed further
in Language Sampling as a Supplemental Measure in Group of Assessments section). The NWR task was strongly correlated with the BESA Spanish, CELF-P2 Spanish standardized assessments, and PLS-4 Spanish \((r = .69, p < .01)\). Therefore, there is no need to administer this task as another language measure in a battery, but it can be useful to examine speech-sound production.

Given these results, the following may be an appropriate battery to differentiate Spanish English-speaking 4-5 year olds with typical language from those with language impairment:

- Parent interview
- A standardized assessment in Spanish
- A standardized assessment in English
- Language sample (If the PLS-4 Spanish is selected as the standardized assessment in the battery.)

It is important to understand that the above battery can be used to identify children with language impairments, but there are other assessment measures that sometimes should be supplemented in a battery because they provide additional information besides for classification. The following assessment measures cannot be used to differentiate TL and LI children, but provide qualitative information that is helpful to describe a child’s specific strengths and weakness when planning for intervention. (The rationale for including each of these assessment measures is provided in other sections of the discussion.)

- Dynamic assessment
- Language sample
- NWR tasks
Language Sampling as a Supplemental Measure in Group of Assessments

As mentioned earlier, a language sample should be administered as an assessment measure in any battery, even a battery containing the BESA Spanish or CELF-P2 Spanish that were strongly correlated with language sample, as predictors of a language impairment. One reason for this is that the language sample provides additional qualitative information that cannot be obtained from the other quantitative variables. In comparison to standardized assessments assessing isolated skills in a decontextualized format, language sampling provides information about the child’s spontaneous language skills in naturalistic contexts so that a clinician can use clinical judgment to assist in the identification of a language impairment. There are also additional pieces of information (besides classification accuracy) that can be analyzed once language sample/s are collected, providing valuable information for the researcher or clinician. First, collecting a language sample in both languages allows the evaluator to assess language proficiency, which can then guide the evaluator to select the other assessment tasks and the specific language for each task. This will ensure that the child will not be assessed only in his/her weaker language (Gutierrez-Clellen, et al., 2000), and the evaluator can tailor the different assessment tasks based on the sociolinguistic characteristics of the child. It should be noted that in this study there was varied performance on production of the language samples. Some of the participants produced a language sample in only one language (English or Spanish), but could not produce a language sample in both languages. This finding highlights the importance of using language sampling to examine language proficiency and is discussed later (in the Participants’ Age and Sociolinguistic Factors in Relationship to Assessment Performance section). Second, language sampling is a non-standardized task that can be collected in a variety of communicative contexts. The evaluator has the flexibility to set up a naturalistic setting for
the child to assess his/her spontaneous utterances using culturally appropriate materials and procedures (e.g., topics). An evaluator can use various elicitation techniques such as play and/or conversation and different conversational partners to elicit the language sample/s. This is advantageous because Latino children’s conversations often focus on family topics and they are more comfortable interacting with peers and siblings instead of adults. Third, the evaluator can gather information about the child’s schema and semantics, in addition to the morphosyntax observed during the production of a language sample. Fourth, an evaluator can compare a child’s language skills by analyzing the MLU-w of the child’s English and Spanish language sample, as well as the frequency of code-switching (Gutierrez-Clellen, et al., 2000). In summary, a language sample serves a dual purpose by providing information to assist a clinician in identifying a language impairment and planning intervention.

As with all other assessment measures recommended for use with bilingual children, there are some problems with language sampling for this population. Language sampling is a static measure that takes a significant amount of time to administer (Stockman, 2010). Since it is a static measure, it examines previous linguistic knowledge and so CLD children might not perform well due to lack of linguistic and/or environmental exposure (Stockman, 2010). However, an evaluator can still benefit from language sampling as long as he/she is aware of the problems so that they use effective elicitation and analysis procedures that were researched with this population. Gutierrez-Clellen, et al. (2000) describe several methodological considerations when using language sampling including accepting a child’s code-switching utterances without penalizing him/her, using appropriate measures for analysis (e.g., MLU-w for bilingual children), comparing a child’s performance to children of similar sociolinguistic backgrounds, and differentiating grammatical errors occurring from the learning of a second language from
those that are characteristic of an English-speaking child with a language impairment. Also, vocabulary size and variety may be due to a child’s limited linguistic and/or environmental exposure and may not effectively discriminate between TL and LI groups. Munoz, Gillam, Pena, and Gulley-Faehnle (2003) found that TNW (total number of words) and NDW (total number of different words) were not sensitive measures for young Latino children from low SES backgrounds and Uccelli and Paez (2007) did a study showing that only NDW was a sensitive measure when analyzing young bilingual Spanish-English speaking children’s narratives. MLU-w and MLU-m were input into the regression model and both showed moderate correlation with language status; however, analysis of the presence/absence of obligatory morphological markers might have indicated a stronger correlation with language status. Gutierrez-Clellen and Simon-Cereijido (2009) specifically studied the use of language sampling with predominantly Spanish-speaking and bilingual Spanish-English children by collecting a language sample in both languages, which provides more information regarding the speech-language skills of participants than a language sample collected in only one language. These authors suggest that besides calculating an MLU for morphemes (MLU-m) using an English narrative language sample and an MLU for words (MLU-w) with a Spanish narrative language sample, a measure of morphosyntactic accuracy should also be measured for each narrative language sample, using the morphosyntactic skills that have been shown to be problematic for each language respectively. The percentage of correct production of finite verbs morphology in English and the percentage of correct production of articles, verbs and clitic pronouns in Spanish should be calculated (Simon-Cereijido & Gutierrez-Clellen, 2009). These authors have stated that a child who performs below 80% on both narrative language samples is likely to have a language impairment. When a language sample in each language is available and a measure of
morphosyntactic accuracy is calculated, a stronger correlation between language sampling and language status might be found.

*Dynamic Assessment Findings*

Previous research has shown that dynamic assessment (DA) tasks discriminate between children from culturally and linguistically diverse backgrounds with language impairment vs. those with limited linguistic/environmental exposure (Kapantzoglou, Restrepo, & Thompson, 2011; Pena, Iglesias, & Lidz, 2001; Pena & Quinn, 1997). It is unclear why the DA task in this study was not a significant discriminator between the two groups, even though current research supports the use of this type of task to assist in the identification of language impairment among CLD children. The same protocol was used in this study as the dynamic assessment task described by Pena, et al. (2001) for word-learning. These authors concluded that the difference in pre and posttest scores along with the modifiability rating scale scores provided the most accurate classification of typical and low language learning ability groups. On the modifiability rating scale, the evaluator records the strategies the child used to learn about naming, while observing during the mediation sessions. Therefore the pretest-teach-posttest design using the DA MLE (mediate learning experience) protocol might be appropriate; however, since only the pretest-posttest performance differences were calculated in this study without the modifiability rating scale, a statistically significant effect was not found. It should be noted that other studies also concluded that use of the modifiability scales was the best method for differentiating between groups of children (Kapantzoglou, et al., 2012; Ukrainetz, et al., 2000). The modifiability rating scale was not used in the current study because it is to some extent a subjective measure.
It might have been a better option to use the DA procedures examining word learning abilities as described in the Kapantzoglou, et al. (2011) study. This study used dynamic assessment with a similar population and ages as the current study, as well as with the same purpose of classifying typical language learners and children with primary language impairment. The authors used a pretest-teach-posttest design to teach three target nonwords with unfamiliar items in a short 30-40 minute session, following the MLE approach described by Pena, et al. (2001). It seems that their procedure was more manageable and focused because it included few nonwords and only one brief session.

**Translated/Adapted Standardized Assessments**

Although the translated/adapted Spanish standardized tests administered in this study differentiated between TL and LI groups well, it is important to note that the individual scores on translated/adapted versions of standardized assessments such as the CELF-P2 Spanish and PLS-4 Spanish might still be inaccurate in regard to a child’s language abilities. This may result from content/linguistic biases of specific items within these tests and flaws in construct validity due to the fact that these tests that were originally developed for mainstream American English-speaking children. A child’s score may not accurately reflect their level of linguistic knowledge in their native language, and the score may not have been a true representation of the child’s language skills/abilities despite the fact that the score placed the child in the correct group. The outcomes of this study showed that these standardized assessments identify children accurately as TL or LI, but actual practice shows that the exact score might still be inaccurate because it is not an accurate representation of the child’s linguistics skills. For example, a child who receives a standard score of 70 on a Spanish standardized assessment is considered to have a language impairment, but he may present with slightly better language skills (and his language functioning
level is really a 78) because the score was negatively influenced by some of the biased items included in the assessment. Then, educators might use the child’s score to make decisions about placement/intervention, and mistakenly place the child in a more restrictive educational environment. Thus, scores should not be reported for clinical purposes. Confidence intervals might be a more accurate measure to be reported.

In summary, the purpose of standardized assessments is only to identify children with language impairments, and the results from this study show that they accurately differentiate TL and LI groups. Standardized assessments can be used to determine the presence or absence of a language impairment in bilingual Spanish-English speaking children; however, since the child’s actual score on a standardized assessment might not represent his/her language abilities, the scores should be interpreted cautiously.

**Language of Assessment for Bilingual Children**

Since bilingual children speak two languages with varied proficiency levels (Bohman, Bedore, Pena, Mendez-Peres, & Gillam, 2010; Goldstein, 2006; Gutierrez-Clellen, et al., 2000) and results from this study showed that assessments in both Spanish and English were predictive of a language impairment, administration of assessment measures in both languages is recommended for this population. The suggested battery includes an English and Spanish standardized assessment, but a clinician still needs to use clinical judgment when deciding the types of tasks that will be used to assess each language. A clinician might administer a language sample in both languages using this information to assess language proficiency and linguistic knowledge/skill, as well as administer English and Spanish standardized assessments and other processing dependent measures (e.g., nonword repetition, dynamic assessment tasks). In another situation where an individual child displays greater Spanish language proficiency and weak
English language skills that cannot be assessed formally, a clinician might select only a Spanish standardized assessment along with other non-standardized assessments (e.g., parent interview, language sample). The results from this study support the administration of a Spanish and English standardized assessment for 4-5 year old bilingual Spanish-English speaking children, but there is no steadfast rule. There is no way to specify exactly which assessments should be given in each language for a group of children because the evaluation of a child needs to be individualized and tailored to assess the individual child’s linguistic knowledge, their strengths and areas of difficulty. This applies especially to the bilingual population where there are many sociolinguistic variables influencing a child’s language proficiency and skill.

*BES A Findings*

The BESA is a new test that was recently developed for the specific population targeted in this study (bilingual Spanish-English speaking children in the United States). It includes Spanish and English subtests based on the cultural and linguistic characteristics of each language.

The analysis for this study included total language measures for all the standardized tests except that the semantics and morpho-syntax subtests for the BESA were also input for analysis, given that it is a new test, and analysis of each subtest in the model could be informative for researchers. The results indicated that based on total test scores, the BESA Spanish assessment may be the most discriminating for this population because it explained most of the variability \( r = -.793, p = .000 \); however, further analysis of the specific subtests indicated that the BESA Spanish morpho-syntax was the subtest that explained the most variability \( r = -.817, p = .000 \) from all of the predictors.

Although the BESA Spanish morpho-syntax subtest was found to be the most discriminating assessment for this population, a clinician would not be advised to give only one subtest to a
child as the child might have difficulty in another language area that would not be assessed by a morpho-syntax subtest. Although the group statistics indicated that this subtest explained most of the variability, an individual child could have typically developing morpho-syntax, but have difficulty in the semantic area. From a practical perspective, a clinician might administer the BESA morphosyntax subtest to collect information without scoring it, but a single subtest would not be used as a basis for clinical decisions as other language areas may contribute to an individual child’s impairment. From a theoretical perspective, it is an interesting finding for researchers that morphosyntax accounts for a large percentage of the variance in a language impairment as it confirms that syntax is a primary area of deficit for bilingual children with language impairments.

The BESA was originally developed for the bilingual Spanish-English speaking children in the United States. Therefore, one might assume that the BESA is better at differentiating Spanish-English speaking TL and LI groups than translated/adapted Spanish versions of English tests that were based primarily on English language characteristics and mainstream American culture. Although this was not a separate research question at the outset of this study, an analysis was able to be completed because the data was already collected and available. A comparison of the standardized tests revealed that the BESA Spanish Total, CELF-P2 Spanish, and PLS-4 Spanish assessments are highly discriminating for TL or LI group membership, but the BESA did not discriminate between groups more accurately than the CELF-P2 and PLS-4 Spanish Editions, which are translated and adapted versions of the English editions.

Although a statistical significant effect was not found between the BESA and other standardized assessments, there may be advantages to using this new test. Since the subtests were created with the specific culture and language experiences for this population in mind, a
child might feel more comfortable with this assessment and respond better. One advantage of
the BESA may be that the different items on the English and Spanish subtests reduce the bias of
administering the same test in two languages. The BESA assessment contains Spanish and
English subtests that include completely different items, as opposed to the PLS-4 and CELF-P2
that are translated so that similar items are included in the English and Spanish versions. A
child’s performance on the second administration of a translated/adapted test might be biased by
the child’s familiarity with the test when it was first administered in another language.

*Standardized Assessment Designed for Testing English and Spanish*

The BESA might eventually prove to be a better standardized assessment because the scores
may be more precise/accurate as items that are not consistent with identifying a language
impairment in the respective languages are not included in the standardized assessment’s
subtests. As mentioned earlier, the BESA (Pena, in development) claims to include items that
are based on the linguistic characteristics and cultural practices of bilingual Spanish-English
speaking children in the United States. It would be prudent to further study whether the BESA
(which was developed for the specific population in this study, assesses both Spanish and
English language skills, and includes different test items between Spanish and English subtests),
provides scores reflecting a child’s true language abilities more accurately than
translated/adapted versions of English tests.

*Participants’ Age and Sociolinguistic Factors in Relationship to Assessment Performance*

The age of the participants and sociolinguistic variables may influence a child’s performance
on Spanish and English assessment measures. The participants in this study were 4-5 years of
age, had less exposure to English than Spanish at the time of data collection, and might not have
yet experienced how social acceptance into the larger community is easier when using the
majority language. Although the results in this study showed that only the Spanish standardized assessments had strong correlations with language status, the English standardized assessments may be more predictive of language impairment as these children become older. Kohnert and Bates (2002) showed that language dominance varies with age, and many Latino children often develop greater English language proficiency than Spanish after a number of years of schooling in the United States. This would also apply to children with language impairment who usually develop greater proficiency in English even though they have lower language skills in both languages compared to age- and experience-matched peers. It is interesting to note that Gutierrez-Clellen and Simon-Cereijido (2007a) found that the BESA English grammatical measure was able to identify LI with fair sensitivity and good specificity among Latino English speaking children; thus, this grammatical measure may prove to be a good predictor as the children in the current study become more proficient in English. To summarize, it might be that English standardized assessments will be good predictors at later ages for this population.

Participants’ performance between the English and Spanish standardized assessments were compared and showed that participants performed higher on the CELF-P2 Spanish than the CELF-P2 English. This suggests that the CELF-P2 Spanish is an easier assessment for bilingual Spanish-English speaking children in the United States (whose native language was Spanish). This finding is reasonable because the children in this study had more exposure to Spanish than English at the time of data collection; however, it is unclear why a significant difference was not found between the PLS-4 and BESA English and Spanish assessments.

Although it was anticipated that all the students would perform better on the Spanish assessment measures because they have been exposed to Spanish for more years than English, other sociolinguistic factors seem to have influenced their performance as there were some
participants who performed higher on English versions of the language sample task.

Specifically, 9 children (5 TL 4 LI) produced both Spanish and English language samples; 7 children (3 TL 4 LI) produced only Spanish language samples; 12 children (6 TL 6 LI) produced only English language samples. These results as well as the literature (Hammer, et al., 2012) suggest that passive exposure to instruction and practice/usage of a specific skill in one language can assist in the development of the skill at a higher level in the targeted language of instruction.

In this study, most of the participants’ exposure to storybook reading and narratives has been at school, where English was the language of instruction. Hammer, et al. (2012) explains that bilingual children’s vocabulary learning and story retelling skills in each language are influenced by both passive exposure in their environment and their use of that language. Studies have shown that when examining bilingual children’s language development, information about exposure and use of language needs to be collected because they both strongly influence the development of the languages (Bohman, et al., 2010; Hammer, et al., 2012). When children in this study were asked to retell a story, the MLU-w varied among children even though all these children were exposed to Spanish for more years than they were exposed to English. As mentioned earlier, some children’s MLU-w was longer in English and some children could produce a narrative only in English. These results suggest that language usage is strongly associated with story retelling skills, which is consistent with the findings from a recent study by Hammer, et al. (2012). Some of these children presented with greater semantic-syntactic abilities in English than Spanish even though they were exposed to English for only one or two years. In addition to exposure to the language of instruction and children’s opportunities to practice storytelling skills in English, the children’s better performance in English might have
also occurred because children frequently learn English quickly after they enter the educational system (Veltman, 1988; Wong Filmore, 1991).

These results allude to the importance of teachers encouraging and facilitating students’ practice in producing and retelling narratives as language use with teachers and peers in an educational setting promotes the learning of the specific language used to tell the story. Transfer between the languages can occur, but the child might demonstrate higher levels of performance initially in the targeted language of instruction.

**Limitations and Future Research**

It should be noted that only one assessment task (in each language) was used for each type of assessment measure except the standardized assessments; thus, these results should be interpreted carefully as the use of a different assessment task could have yielded different outcomes. For example, the Restrepo (1998) parent interview was used in this study, but the BESA parent interview, which was recently developed and collects detailed information about the amount of exposure to each language, might indicate a different relationship between parent interview and group membership if it was used instead of the Restrepo (1998) parent interview. A more detailed parent and teacher interview examining the amount of exposure to English and Spanish and the child’s use of each language is also recommended for use in future studies, based on the results from this and other studies (Bohman, et al., 2010; Hammer, et al., 2012) that showed the influence of exposure and usage of each language on language proficiency and skill in that language.

The dynamic assessment task was the only assessment that was administered only in English or Spanish as it was not feasible to use the same Mediated Learning Experience (MLE) script described by Pena, et al. (2001) to teach the same words first in one language and then the
other because of the biases that would have existed. The data collectors were encouraged to use language (English or Spanish) that matched the child’s language usage and to provide individualized feedback to participants’ responses. The lack of a standard protocol may be related to the fact that this assessment was not a statistically significant predictor, but has been shown to be useful in other studies. It is suggested that further studies include a variety of dynamic assessment tasks using a pretest-teach-posttest design, along with the modifiability rating scale (Lidz & Pena, 1996).

Many researchers and bilingual speech-language pathologists criticize the reporting of actual scores when administering a translated/adapted version of an English test because the scores might not represent the child’s language skills. It would be prudent to further study whether the BESA provides scores describing a child’s language abilities more accurately than translated/adapted versions of English tests.

The results showed an interesting theoretical finding for researchers that the Spanish morpho-syntax subtest was a major indication of language impairment. From a theoretical perspective, researchers might follow-up to further understand the degree that morpho-syntax accounts for a language impairment in children compared to other language areas assessed (e.g., semantics, NWR).

Since morpho-syntax is a strong indicator of language impairment, perhaps morphology should be examined further with some of the assessment measures included in this study. A dynamic assessment task examining a child’s ability to acquire morphological skills through a test-teach-retest design might be useful for identifying language impairment. Also, MLU-m and MLU-w calculations for the language samples showed moderate correlations with language
status; however, future studies that examine the presence/absence of morphological markers in obligatory contexts might result in stronger correlations with language status.

Since many of the children could not produce language samples in both English and Spanish, it would be useful to examine whether eliciting language samples using a play instead of a story retelling context would yield a greater number of language samples in both languages. The children might have an easier time using both languages for social (play) tasks instead of academic (story retelling) tasks. A future study can also investigate the usefulness of language sampling for intervention planning since the primary purpose of the current study was to examine language sampling only for the identification of language impairment.

In order to externally validate the battery recommended in this study, the battery should be applied to another sample (of children who are not yet identified as having a language impairment). It would also be beneficial to administer all of the assessment measures to a different group of identified children to confirm the results from this study. Last, the sample size was small because of the increased number of predictors. A follow-up research study with a larger sample size to achieve sufficient statistical power for the number of predictors is recommended.

Implications

This study strengthens the current body of literature on bilingual assessment measures by indicating that all of the current assessments except the dynamic assessment and NWR English tasks previously suggested for use with this population are valuable in the identification of a language impairment. This study confirms previous findings that examined the usefulness of these assessment measures, as well as supports their use with 4-5 year old bilingual Spanish-English speaking children in the United States. Furthermore, a battery containing the fewest
assessment measures with the highest level of confidence was determined to identify children with language impairment effectively and efficiently. Based on the data, the battery should include Spanish and English standardized assessments, a parent interview, and a language sample (when the PLS-4 Spanish is included as the Spanish standardized assessment). The discriminant model showed that the battery has a sensitivity value of 86.6% and specificity of 93.3%. These values are above the minimum established by Plante and Vance (1994) indicating that the battery suggested in this study is appropriate to identify a LI in 4-5 year old bilingual Spanish-English speaking children accurately. Based on clinical judgment, there are other assessment measures that sometimes should be supplemented in a battery because they provide additional information to describe a child’s specific strengths and weakness when planning for interventions. These include dynamic assessment, language sample, and NWR tasks. As mentioned earlier, the purpose of this study was to determine those assessment tasks appropriate for the identification of a language impairment; however, clinicians usually make decisions regarding identification and intervention planning (frequency, duration, goals) simultaneously within one evaluation. Therefore, although the statistical results showed that the language sample cannot be used alone for identification, it is a useful supplemental measure that assists in both identification and intervention planning to describe a child’s current language skills and intervention targets.
References


