2017-01-01

Validating The Accuracy Of A Language Screening Instrument In Identifying Preschool Children In El Paso, Texas

Roxana Stubbemann
University of Texas at El Paso, roxanastubbemann@gmail.com

VALIDATING THE ACCURACY OF A LANGUAGE SCREENING INSTRUMENT IN IDENTIFYING PRESCHOOL CHILDREN IN EL PASO, TEXAS

ROXANA STUBBEMANN

Master's Program in Speech-Language Pathology

APPROVED:

Connie Summers, Ph.D., CCC-SLP Chair

Vanesa Smith, M.S., CCC-SLP Member

Carla Contemori, Ph.D Member

Charles Ambler, Ph.D.
Dean of the Graduate School

Dedication

I would like to thank my family, professors, and friends for always supporting my education and academic endeavors. They have provided continuous love and support throughout my academic career and have always encouraged me to pursue my dreams in life. They encourage me to be the best clinician and researcher I could be and believe in my capability to make a change in the world.

VALIDATING THE ACCURACY OF LANGUAGE SCREENING INSTRUMENTS IN IDENTIFYING PRESCHOOL CHILDREN IN EL PASO, TEXAS

by

ROXANA STUBBEMANN

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Program in Speech-Language Pathology

THE UNIVERSITY OF TEXAS AT EL PASO

May 2017

Acknowledgements

Foremost, I would like to express my gratitude to my advisor Dr. Connie Summers for her guidance, mentorship, and valuable feedback as I wrote this thesis. I also thank Mrs. Vanesa Smith for her time contributed in preparing and carrying out this research study. Additionally, I would like to thank the graduate and undergraduate students in the UTEP Research in Bilingual Language Learning Lab for their time and dedication towards collecting, transcribing, and coding data. Without their assistance and support, this thesis would not be possible.

Abstract

Background: Effective screeners are time and cost effective and may be an efficient indicator of whether a child needs further in-depth assessment. Surveyed Speech-Language Pathologists in the El Paso, TX region report using informal screening tools such as observational measures and checklists due to the lack of adequate standardized tools that suit the needs of bilingual children, which occupy a majority of their caseloads (Curtis, Summers, Smith, & Stubbemann, 2016). Many language-screening instruments are not psychometrically sound for culturally and linguistically diverse (CLD) populations and limited research exists on the accuracy of these tools.

Purpose: The purpose of this study is to expand the limited research of whether a proposed instrument, the Bilingual English Spanish Oral Screener (BESOS) accurately identifies CLD preschool-aged children who are at risk for language impairment (LI) and is valid for use in the El Paso, TX region.

Methods: Nineteen children, aged 3-6 years, participated. Participants were screened for LI using the BESOS screening instrument, and assessed using a gold standard (PLS-5 and Language Sample). The sensitivity and specificity of the BESOS were gathered to determine the accuracy in identifying preschool children for the El Paso, TX region.

Results: Results indicated an overall sensitivity of 100% and specificity of 81%. The BESOS had a good sensitivity (100%) in accurately identifying "at risk" participants for LI in all groups. For Pre-School however, specificity was (40%). Specificity for Pre-Kinder and Kinder groups were inconclusive as no children with LI were present.

Conclusion: Preliminary data of this study shows promise for the use of the BESOS for identifying children at risk for LI in El Paso, TX.

Table of Contents

List of Tables

Chapter 1: Literature Review

1.1 Introduction

Differentiating language differences from a true impairment is a difficulty many Speech-Language Pathologists encounter when working with individuals from culturally and linguistically diverse (CLD) backgrounds. Language impairment (LI) is prevalent in 7.4% of the English-monolingual population (Tomblin, Records, Buckwater, Zhang, Smith, & O'Brien, 1997). However, it is much more difficult to identify children when multiple languages are influencing one another. Identifying LI in individuals learning English as a second language is difficult because English-language skills are often still developing and monolingual normed assessments are not sufficient for bilinguals (Paradis, Schneider, & Duncan, 2013). Signs of LI emerge in all languages during pre-school and school age years. Impairment may additionally influence any or all of the five language domains: phonology, morphology, semantics, syntax, and pragmatics. Children with LI may continue to experience difficulties in academia due to literacy difficulties, behavioral challenges, and mental health problems (Prelock, Hutchins, & Glascoe, 2008). Thus, early identification of impairment in the pre-school aged years is crucial to the early intervention process.

A specific tool to assess pre-school children for LI is a language-screening instrument, also known as a "screener". Language screeners identify those at risk for impairment and indicate whether further, in depth, assessment is necessary (Guiberson & Rodriguez, 2010). Many language screeners, however, are not culturally sensitive, valid, nor reliable. The majority of commercially available screeners are normed on monolingual children, which may lead to a disproportionate number of children failing a screener. In these cases, a language difference is presenting itself instead of LI (Paradis, Schneider, & Duncan, 2013).

The need for psychometrically sound language screeners has been a reoccurring issue that researchers have been investigating for decades, in both monolingual and bilingual measures. Law, Boyle, Harris, Harkness, and Nye (2000) conducted a systematic review of literature focused on the sensitivity, specificity, and likelihood ratios of various authors and their testing of screening tools. This review found that specificity was higher than sensitivity across different instruments suggesting it is easier to accurately identify children who do not have LI from those who do have LI. With 81.2% of the El Paso population reporting as Hispanic or Latino and with 73% of the population speaking a language other than English at home, valid screening practices are an essential issue for this region (U.S. Census Bureau, 2014).

1.2 Current Screening Practices in El Paso, Texas

To address screening practices in El Paso, TX, a survey was conducted to gain insights into the perceptions of professionals on the effectiveness, as well as the types of screening instruments being used in the region (Curtis, Summers, Smith, & Stubbemann, 2016). Due to the demographics of the area, examining the screening practices, as well as the efficacy of such screening practices, is a great need for the community. Professionals in the area were surveyed to address local needs in screening.

Out of 43 speech-language pathologists surveyed, 77% reported using screening measures with English-Spanish bilinguals (Curtis, Summers, Smith & Stubbemann, 2016). The speech-language pathologists who reported screening children, overwhelmingly used informal screening measures. Eighty-five percent reported using observation; 60% teacher checklists; and 40% parent checklists as measures to screen children. These methods were preferred over formal assessments. When questioned about weaknesses in their screening practices, 54.8% of participants responded with a theme of not having the appropriate tools to screen. In addition,

38.7% responded that they did not feel screeners were accurate nor consistent tools. Overall, data suggested that speech-language pathologists in the El Paso area might have lacked confidence when using formal screening measures to screen bilinguals (Curtis, Summers, Smith, & Stubbemann 2016; Guiberson & Rodriguez 2014). Lack of confidence may have been due to inadequate training in interpreting results for bilingual populations or difficulty looking at language holistically (i.e. not as independent languages). Survey findings suggested the need for valid, reliable and accurate measures to better contribute to the efficacy of clinical practice in the El Paso, Texas region.

1.3 Advantages of Screening Instruments

Screening instruments are designed to determine "at risk" children. In doing so, many children are over-identified with LI. Over-identification will cause children to fail the screener even though impairment may not exist (false negatives) (Allen & Bliss 1987). By screeners over-identifying children, it ensures that the children who need further assessment get the services they need, without going undetected. Individuals who do not really have a deficit will be released upon further assessment.

Screeners are also important because they are an efficient use of time and money. Language development and growth are most rapid in the early years of life. Therefore, early identification through screening may increase prognosis for treatment success. (Allen & Bliss, 1987; Eriksson Guiberson & Rodriguez 2010, Texas Department of Assistive and Rehabilitative Services, 2014; Westerlund, & Miniscalco 2010). As a precursor to a full assessment, a screener may indicate the need for a more in depth examination of a child's language abilities (Allen & Bliss, 1987). Although a screening instrument may be used as an alternative to a full assessment initially, a screener does not replace a full assessment. Standardized assessment administration

can be time-consuming during a full evaluation (i.e. 2-3 hours) and can be costly in materials. Screeners are advantageous because they are quick and give a brief synopsis into the child's language abilities while assessments are labor-intensive and financially diminish resources, which would be better allocated to individuals who require and could benefit from intervention. (Allen & Bliss, 1987).

The use of screeners is especially advantageous for those who work in the schools and acquire large workloads. In the U.S., the average caseload size for a Speech-Language Pathologist working in the schools is 48.8 individuals (Katz, Fallon, Blenkarn, & Smith 2010). In 1993, the recommended caseload size, as per the American Speech-Language-Hearing Association (ASHA), was 40 individuals (Katz, Fallon, Blenkarn, & Smith 2010). In addition to caseload, workload includes evaluation and assessment time, which may be reduced by screeners. Katz (2010) found that those with smaller workloads had higher job satisfaction than those with larger workloads, which is an additional advantage of screeners. Accurate screening tools may be efficient in that they reduce the amount of time spent on full assessments, as well as act as a good predictor of performance with a much shorter administration time (20-30 minutes).

1.4 Challenges in Screening Instruments

As many screeners are not adept to CLD populations, many issues arise when assessing populations who speak or understand more than one language (Bedore & Peña 2008; Brebner et al., 2015). There are few screeners that take second languages into consideration and thus, a child may be screened in one language and fail when compared to the standardized scores of an individual who is monolingual. Features of the language may interfere with a true representation of the child's language skill. Some issues include the individual's proficiency in each of the languages, the fact that language development differs for bilingual children, and the acquisition

for both languages is influenced by exposure to language and to the content. (Bernal & Tucker 1981; Brebner et al. 2015; Peña et al. 2012).

When screening instruments use a normative sample that does not fit an individual child's profile, results will be language biased and non-representative of the individual. Often, screeners do not account for variations among languages. For example, if a picture of a stimulus item is presented and the child responds in English but uses the syntactic structure of Spanish, the item would be scored as incorrect because the test is examining English only. This example would account for differences between languages, not necessarily LI. According to ASHA (2016), in order for a child to be identified with impairment, the impairment must be evident in all languages used by an individual.

1.4.1 English language learners

Because all languages must be affected for an impairment to exist, assessing each independent language is necessary to acquire a representation of the child's true abilities. Bilinguals may appear to have lower vocabulary than their monolingual peers, however, lower vocabulary may not be a result of a handicap but perhaps a smaller variety of language input taken from each language separately (Baetens-Beardsmore, 1986). Individual's who learn English as a Second Language will comprehend more than they can express in English, while being proficient in their home language (Bedore et al., 2012). This may undermine their language capabilities and mimic the presence of a LI if English alone, was examined.

1.4.2 Cultural biases of screening instruments. Contributing factors to the inaccuracy of screening tools include cultural biases that may underestimate a child's abilities. One such bias stems from the fact that most screening instruments have not been standardized on children from different "cultural, economic, or linguistic backgrounds" (Norris, Juarez & Perkins 1989).

Other biases related to exposures of an individual include biases related to language, societal values, religion, or life-experiences.

Cultural, economic, and linguistic factors may be seen through different biases such as examiner bias, item bias, or method bias. For example, Norris, Juarez, and Perkins (1989) found that individuals performed better during assessment when the examiner was of the same cultural background than a different cultural background. Such results show that there is a potential examiner bias that may exist with children from CLD backgrounds. Item bias can occur when children from low SES and cultural minority backgrounds score lower on standardized measures due to the lack of life-experiences and lack of exposure to some of the pictured stimulus items (Hoff & Tian, 2005). Method biases can occur when participant responses are measured in different ways. For example, when looking at the language of a Spanish speaker, utterances may appear shorter due to a difference in sentence structure than in English. Speech-language pathologists must recognize the influence of culture on individual performance as well as the need for assessments that take such biases into consideration.

1.5 Addressing Cultural Gaps in Assessment

Researchers and community members have acknowledged the importance of having culturally and linguistically appropriate considerations. In an attempt to suppress cultural biases, many methods of adaptation towards assessments and screening instruments have and are continuing to be made (Camilleri, Hasson, & Dodson, 2014;Geisinger, 1994; Gretch & Dodd, 2007; Gross, Buac, & Kaushanskaya, 2014; Gutierrez-Clellen & Peña, 2001; Herdman, Fox-Rushby, & Badia, 1997; Lugo-Neris, Peña, Bedore, & Gillam, 2015; Peña, 2007; Patterson et al., 2013). Attempts include a continuum of efforts from the translation of materials to using different norming samples (Geisinger, 1994). Although these contributions have added to

improved screening practices for CLD populations, translations of materials carry many flaws. When new tests are developed, they must be validated to assure their diagnostic accuracy.

As professionals have become more aware of the dangers of directly translating forms and documents, recent research has begun to look at more functional approaches to this problem. It has been suggested that formal norm-referenced tests be combined with informal tests, developmental checklists, and clinical judgment in order to engage in the best practice for screening bilingual populations (Lugo-Neris, Peña, Bedore, & Gillam, 2015). There is never perfect accuracy of any given instrument so a combination of them is always recommended. A clinical decision should never be made based on a single instrument.

1.5.1 Dynamic assessment. In combination with using multiple instruments and tools, researchers have also addressed the concern of testing bilingual populations through methods such as dynamic language assessment, which assess learning potential in those from CLD backgrounds (Camilleri, Hasson, & Dodson, 2014; Gutierrez-Clellen & Peña, 2001; Patterson et al., 2013). Dynamic assessment techniques include a pre-testing period, teaching period, and post-testing period. The focus and rationale of dynamic assessment is to identify if a different learning experience or lack of educational opportunity is existent rather than impairment (Gutierrez-Clellen & Peña, 2001). Because it is often difficult to differentiate a language difference from impairment, dynamic assessment allows for a clinician to gather insight to difficulties noted on the surface.

Dynamic assessment has also been used as a screening method described by Patterson and Dale (2013). Authors suggested incorporating a short teaching period as part of the screening process to see if children have the capability of learning the material. Dynamic assessments are used to determine if difficulties are due to impairment or if difficulties are due to

lack of exposure (i.e. insufficient language model, limited access to language mediums) in the parameter being tested.

Due to the added opportunity to grasp concepts during teaching periods, lack of exposure will result in the data showing a high number of prompting for beginning tasks and a lower number of prompts for later items. If dynamic assessment were used for screening, the speech-language pathologist would have greater insight into the root of difficulties (language difference or LI) and would allow for the best clinical decisions if the screener results in a failing score.

This screening method proposed is an alternate method often used for an effective assessment of individuals of a CLD background but also has some disadvantages. Patterson and Dale (2013) found that the time it took to dynamically screen was much longer (30-45 minutes) than compared to what an average screening tool typically takes. Shorter administration time often makes screeners appealing to clinicians. Long administration time found in the study is also problematic because those screened were all typically developing children and screening a child with LI, may actually increase the duration of the screening.

1.5.2 Conceptual scoring. Another in-work strategy for bilingual screening instruments is combining scores from the two languages being assessed. Conceptual scoring is scoring the meaning of the response an individual gives, regardless of the language of response versus focusing on the word (Pearson, Fernandez, & Oller, 1993). Sheng, Peña, Bedore, and Fiestas (2012) revealed that using a conceptual score resulted in a higher score than did either single-language score (Gross, Buac, & Kaushanskaya, 2014; Sheng, Peña, Bedore, & Fiestas, 2012). Conceptual scoring allows the examiner greater insight into the individual's language and provides a holistic viewpoint of language development and abilities. Conceptual scoring also allows for a type of compensation when language exposure may be a factor that may not

otherwise be allowed if language scores were viewed independently (Pearson et al., 1993; Sheng, Peña, Bedore, & Fiestas, 2012; Peña, 2015). Children with LI have difficulties with processing and semantic representations (Sheng, Peña, Bedore, & Fiestas 2012). Language impairment influences all languages of a child, so considering that their knowledge is spread between all of their languages allows for more accurate and better informed decision making during the diagnostic process. Conceptual scoring can be incorporated into assessments that test more than one language as seen in many already developed tools as well as tools currently under development or in preparation.

1.6 The Bilingual English Spanish Oral Screener (BESOS).

A specific standardized tool currently in preparation that attempts to use strategies to minimize the cultural gap seen in screening measures is the Bilingual English Spanish Oral Screener (BESOS; Peña, Bedore, Iglesias, Gutiérrez-Clellen, & Goldstein, in preparation). The BESOS is an instrument developed specifically for bilingual children, with a norming sample containing both monolingual Spanish and English children in addition to English-Spanish bilingual children. Additionally, the BESOS uses conceptual scoring to account for language differences. Norming sample features and conceptual scoring are important in that few other language-screening instruments are currently norm-referenced in Spanish (Lugo-Neris, Peña, Bedore, & Gillam, 2015). Other Spanish instruments, however, use Spanish dominant participants in their norming samples, unlike the BESOS, which includes varying levels of Spanish and English proficiencies.

Studies focusing on the development and accuracy of the BESOS have revealed that BESOS subtests (Semantics and Morphosyntax) are highly correlated with each other (between .64 and .87). The most accurate sensitivity of the test is during ages 4;6 and 5;6 (Peña, Gillam,

Bedore, & Bohman, 2011; Summers, Bohman, Gilliam, Peña, & Bedore, 2010). Measures with varying bilingual norms benefit regions such as El Paso, Texas due to high prevalence of bilingualism. Administration of the BESOS occurs in both languages for each of its subtests (Semantics and Morphosyntax). The screener also accounts for responses in another language through the use of conceptual scoring to ensure language differences are noted, rather than language deficits.

The BESOS has been developed as an attempt to address many cultural and linguistic biases. Peña and Bedore (2011) found that classification accuracy for the BESOS' proceeding full assessment, the Bilingual English Spanish Assessment (BESA) was dependent on languages used for analysis and that use of a best score for each subtest resulted in improved specificity. Lugo-Nerris et al. (2015), found that predictive accuracy was validated after a two-year time frame. Lugo-Nerris et al. (2015) stated, "By selecting their best score in each domain, we maximized the likelihood of accurately capturing their ability and attempt to control for some of these differential experiences" (pg.434). Because of the appropriateness of the norming sample and developmental design of the BESOS for bilingual children, the diagnostic accuracy of the Bilingual English Oral Screener will be investigated and validated through the preliminary data in this pilot research study.

1.7 Motivation for Study

There is a lack of research in the area of screener accuracy and even less in bilingual populations. Due to lack of research in bilingual screening, many speech-language pathologists and other professionals who administer screeners may lack confidence in their tools as a result of their uncertainty in its accuracy (Guiberson & Rodriguez, 2010). Currently, there is not sufficient data of well-functioning measures. The current study can be influential in providing

the diagnostic accuracy of a specific tool (Law, Boyle, Harris, Harkness, & Nye 1997). Knowledge of diagnostic accuracy will allow for greater effective, efficacious, and efficient practice as well as provide information to support the use or disuse of certain instruments that aid in making diagnoses for bilingual children.

The proposed screening instrument, the BESOS, is aimed to increase efficiency of screening practices for bilingual children, making it a potentially useful tool in El Paso, Texas. Speech-language pathologists are not uniformly using any one screening instrument over another. A preliminary investigation exploring the validation of the accuracy of the BESOS may serve a local need for the community, as it has been developed under similar target populations.

1.8 Research Questions

The limited research on the accuracy of screeners in bilingual pre-school children has led to the current pilot study. Based on what is currently known about El Paso screening practices and the lack of available tools, the current study takes a deeper look at the following questions:

1. What is the sensitivity and specificity of the BESOS in the El Paso area in identifying pre-school children who are at risk for language impairment from those who are not at risk?
2. Are there differences in the accuracy of the BESOS in identifying Pre-School, Pre-Kinder, and Kindergarten aged children?

The author's hypothesis to the research question:

- The BESOS screening test will be accurate in identifying bilingual preschool children at risk for language impairment.
- There will not be differences in accuracy of the BESOS in identifying at risk children between groups.

Chapter 2: Methodology

2.1 Participants

2.1.1 Recruitment. Nineteen children were recruited from a Spanish-English bilingual pre-school program located in a central region of El Paso, TX. The facility was given a letter of purpose, explaining what the study entailed and what was required of their facility (time, space, setting, etc.) during approval of partnership between facility and research project. As part of the recruitment process, the pre-school was offered an in-service workshop for faculty and staff to compensate for use of facility and recruitment of participants in the setting. In addition, all families who participated in the study received a $40 Albertsons gift-card funded by the University of Texas at El Paso Graduate School Dodson Award for their participation in the study.

2.1.2 Criteria. Inclusionary criteria included ages 3-6, and exposure to both English and Spanish. Exclusionary criteria included having any form of hearing loss or non-verbalism. In addition, participants were required to have written informed consent by legal guardians and a fifteen-minute informational meeting.

2.1.3 Informed consent. The legal guardian for each of the participants in the study was made aware of the process through a notice sent home with the children by the school with information about confidentiality, procedures, and the potential risks and benefits of the study. Potential risks included pullout time during class and fatigue. Benefits of the study included information regarding the participant's language performance on the assessments. Parents were additionally notified that the research was conducted by the University of Texas at El Paso and that treatment would not be given on part of the examiners.

2.1.4 Participants. Of the 19 participants recruited, 10 were males and 9 were females. Ages of participants ranged from 3;2 to 5;11 years and all recruited participants met the inclusionary and exclusionary criteria of the study. Participants were divided into 3 groups; Pre-School, Pre-Kindergarten, and Kindergarten (see Table 2.1.1). Participants included 8 Pre-School children (3;2- 3;11), 7 Pre-Kindergarten children (4;1- 5;1), and 4 Kindergarten children

(4;11-5;11). Dominance for each participant was also determined. To determine dominance, the language with the greatest output score on the Bilingual Input Output Survey (BIOS) was noted. Parent reported home output levels included 13 English dominant children, 5 Spanish dominant, and 1 French dominant. Teacher reported school output levels on the other hand included 13 English dominant, 4 Spanish dominant, and 2 equal bilinguals (see Table 2.1.2).

Participants were additionally deemed monolingual or bilingual based on information provided through parent and teacher interviews and questionnaires. Although the sample size for this study is small, the population gathered in this pilot study will contribute to a larger study, which will be better representative of the El Paso area in terms of language use.

Table 2.1.1
Mean reported ages

	n	***Mean age in months***	***Standard Deviation***
Pre-school	8	43.63	4.46
Pre-kinder	6	50.5	1.53
Kinder	5	64	5.57

Note. Pre-K= Pre-Kindergarten, K= Kindergarten

2.1.5 Parent and teacher questionnaires. Parents and teachers were required to complete two questionnaires; the BIOS and Inventory to Assess Language Knowledge (ITALK), which were pulled from the Bilingual English Spanish Assessment (BESA; Peña et al., 2014). The questionnaires explored history of English and Spanish use at home and school as well as the child's exposure to each of the languages through parent and teacher report.

The Bilingual Input Output Survey (BIOS) included information based on input and output levels in English and Spanish, which provided examiner insight in determining language exposure and dominance, as mentioned above. The BIOS included questions for caregivers and teachers. The survey focused on exposure to languages by outlining a schedule of the child's day such as; whom they are with, and what languages are spoken at those times to gather input and output percentages. A child, for example, with an input report of 40% was spoken to in that language 40% of the time. Additionally, a child with an output report of 40% verbally spoke that

language 40% of the time. Input and output levels gathered through the BIOS are shown in Table 1 (Peña et al., 2014).

Table 2.1.2
Mean reported percentages of input and output

BIOS	***Parent***		***Teacher***	
	Input	***Output***	***Input***	***Output***
Preschool				
English	40.4%	48.15%	50%	45.31%
Spanish	59.6%	51.85%	50%	54.69%
Pre-Kinder				
English	66.6%*	88%*	54.83%	77.58%
Spanish	33.4%*	12%*	45.17%	21.42%
Kinder				
English	59.8%	79.6%	50%	83.4%
Spanish	40.2%	20.4%	50%	16.6%

Note. * One participant in the Pre-Kinder group was excluded from the mean due to French input and output in the home.

2.2 Screening Instrumentation

2.2.1 Bilingual English Spanish Oral Screener (BESOS). The Bilingual English Spanish Oral Screener (BESOS) (Peña, Bedore, Iglesias, Gutiérrez-Clellen, & Goldstein, in preparation) is a screening instrument, not yet commercially available, designed to determine if a child is at risk for LI. The BESOS is targeted towards bilingual children and includes two subtests (i.e. Semantics, Morphosyntax) for each of the languages in order to have a comprehensive assessment of the child's language. Aspects of language that the BESOS investigates include third person singular, auxiliary negation, passives, sentence repetition, vocabulary, categorization, and comprehension (Peña et. al, in preparation). The child is presented with items as they look at digital pictures, via iPad. The BESOS uses binary scoring, 1- correct, 0- incorrect. Also available to report were "NR"- no response and "OL"- other

language to account for language differences, in all form of responses (Peña et. al, in preparation).

Raw scores of the BESOS were converted to Z-scores in which a standard deviation of -1 was used to determine if the participant "failed" each subsection of the screener. A "Best Language Score" was then derived after determining the language the participant performed best in, for each subtest (Peña et. al, In progress). For example, if a participant received a semantic z-score of -.5 for English and a -1 for Spanish, their English score would be considered their "best language score". For the purpose of this study, a <-1 SD for one out of the two domains resulted in a fail or "at risk" and > -1 resulted in a pass or "not at risk" as based off of a study by Lugo-Neris, Peña, Bedore, and Gilliam (2015). In the current study, if a participant failed either the Semantic or Morphosyntactic subtest in their best language, the participant was classified as failing the screener.

2.2 Gold Standard Instrumentation

2.2.1 Preschool Language Scales-5th Edition (PLS-5). The Preschool-Language Scales fifth edition (PLS-5) is a comprehensive assessment instrument used to examine a child's expressive and receptive language (Zimmerman Steiner, & Pond, 2011). The PLS-5 includes receptive (Auditory Comprehension) and expressive (Expressive Communication) subtests with various items on attention, play, gesture, vocal development, social communication, semantics, language structure, integrative language skills, and emergent literacy skills such as book handling or school readiness skills (Zimmerman, Steiner, & Pond, 2011). Participants in the current study were given the English or Dual version of the PLS-5, based on output levels in each language. The PLS-5 required an easel picture book and manipulatives to assess the child's language. Raw scores for the PLS-5 were converted to Standard Scores for each of the subtests

following scoring guidelines including a Total Language Scores (Zimmerman, Steiner, & Pond, 2011). A Total Language Score <-1.5 standard deviation was used to determine a "fail" in the assessment.

2.2.2 Language Sample Analysis (LSA). A picture description task using procedures adapted from Eisenberg and Yu Guo (2015) was used for a Language Sample Analysis (LSA). Participants were provided with 7 pictures and were instructed to select which they wanted to talk about first. The child was then asked to tell the clinician about the picture, and was prompted through a hierarchy of four pre-selected questions to elicit the language sample (Eisenberg & Yu Guo, 2015). Open-ended prompts allowed clinicians to lengthen the sample and assure a good representation of language abilities was gathered, as well as to build rapport and account for any apprehensiveness or anxiety.

Participants were given LSA in both English and Spanish. However, administration of one language (least dominant) was discontinued if deemed appropriate by clinician. Reason for discontinued administration includes production output of alternate language instead of target language, after given cues and prompts to use specific language during presentation of the first picture. Using SALT software, trained research assistants transcribed and coded language samples (Miller, Andriacchi, & Nockerts, 2011). Following protocol of Eisenberg and Yu Guo (2015), a cut-off of 58.38% grammatical utterances was used to determine if language deficits persisted, resulting in a "pass" or "fail" of the assessment. Pictures in the current study were not those used in Eisenburg and Yu Guo (2015).

2.2.3 Bilingual English Spanish Assessment Questionnaire (ITALK).

The ITALK provided questions in relation to proficiency in vocabulary, speech sentence production, grammaticality, and comprehension on a 5-point Likert-scale for each of the

languages (Peña et al., 2014). The survey is aimed for parents and teachers to assess proficiencies in each of the languages at school and home. The highest scores of the two languages were used to determine if concern by caregivers was apparent. According to the protocol, if the highest rating of the two languages was greater or equal to 4.18, no concerns were apparent (Peña et al., 2014). The questionnaires allowed for insights into areas of concern on behalf of parents and teachers as well as to provide a comprehensive insight into performance levels outside the study's sessions, as rated by caretakers (Peña et al., 2014). Parent and teacher proficiency ratings gathered through the ITALK are shown in Table 3.2.3 in the results section.

2.3 Research Design

A random block design was implemented in this study to control for a potential validity threat due to the sequence of presentation for each of the instrumentation tools used. Saville and Wood (1991) stated, "This enables us to account for some of the variation between experimental units and to see more clearly the variation between treatments" (pg.299). Table 2.3.1 and Table 2.3.2 show the different sequences of stimuli that the participants were exposed to.

Table 2.3.1
Bilingual sequence of instrumentation

Independent Variable	***Sequence I***	***Sequence II***	***Sequence III***	***Sequence IV***
Blocking Variable	BESOS (S) LSA (S) BESOS (E) LSA (E) PLS-5 Dual	BESOS (E) LSA (E) BESOS (S) LSA (S) PLS-5 Dual	PLS-5 Dual BESOS (S) LSA (S) BESOS (E) LSA (E)	PLS-5 Dual BESOS (E) LSA (E) BESOS (S) LSA (S)
# of Pre-school	3	4	1	0
# of Pre-Kinder	0	0	2	1
# of Kinder	0	0	1	1
Total:	3	4	4	2

Note. BESOS= Bilingual English Spanish Oral Screener, LSA= Language Sample Analysis, P-5= Pre-school Language Scales-5th edition, (S)= Spanish, (E)= English

Table 2.3.2
Monolingual sequence of instrumentation

Independent Variable	***Sequence I***	***Sequence II***
Blocking Variable	BESOS LSA PLS-5	PLS-5 BESOS LSA
# of Pre-school	0	0
# of Pre-Kinder	2	1
# of Kinder	3	0
Total:	5	1

Note. BESOS= Bilingual English Spanish Oral Screener, LSA= Language Sample Analysis, PLS-5= Pre-school Language Scales-5th edition

Out of the 19 possible participants, children were randomly assigned to each of the six sequence groups; dependent on language output levels determined by the BIOS and ITALK (Peña, 2014). In creating randomized blocks, the BESOS English and BESOS Spanish were not administered sequentially to avoid test influence. Participants were tested over a course of five testing days to prevent fatigue and ensure minimal pullout time from classroom activities.

2.4 Experimental Procedure

2.4.1 Setting. All assessments were conducted once a week on site of the pre-schools facility. A large and unoccupied room was reserved for assessment, containing three large tables and chairs. Up to three children were pulled out of class at a time by Master's level graduate students and tested while under direct supervision of a licensed speech-language pathologist, accredited by the American Speech-Language-Hearing Association. Trained undergraduate research assistants additionally were present to aid in instrument administration and redirection of participants.

2.4.2 Hearing screening. Bilateral pure tone hearing screenings were conducted in accordance to ASHA standards of 20 dB at the frequencies, 1000, 2000, and 4,000 hertz for each of the participants using a portable air-conduction audiometer. The stimuli were presented in one ear at a time and in ascending order. The participant was asked to raise their hand that

corresponded with the ear in which the tone was heard. All participants passed hearing screenings.

2.4.3 Administration of assessments. Using the random block research design, the BESOS, PLS-5, and LSA were administered after being randomly assigned to an order. This design was done to eliminate any biases that may have occurred due to exposure of one test before administering the other. A random research design also attempted to eliminate unwanted effects such as fatigue as a side-effect of one test, a single test influencing the other test(s) and thus, influencing the data collected from the administrations. For children who were Spanish-English bilingual, the tests were randomly administered in both languages. The examiner was positioned so that the child was unable to look at what the examiner was writing down on the scoring sheet. Additionally, non-specific verbal praise was given. Examples of responses to be given include the following: "I like the way you are sitting!", "I love the way you are doing what I ask of you", "You're behaving so well", etc. Prompts to keep going in moments of frustration and non-cooperation were given as needed (i.e. stickers awarded for good behavior). For each of the tests that were administered, the protocol was followed as written. However, fatigue and time constraints were considered. If participants expressed or demonstrated fatigue, administration of assessments were discontinued for the day and resumed the following week.

2.5 Data Analysis

2.5.1 Reliability. To determine diagnostic accuracy in this study, the combination of the LSA, ITALK, and PLS-5 scores are considered the gold standard to validate the BESOS. Inter-rater reliability was completed for scoring of protocols (i.e. PLS-5, BESOS) as well as data entry of those scores into an excel worksheet. A single transcriber coded 21% of English samples and 25% of Spanish samples independently. Additionally, a second trained transcriber reviewed LSA transcriptions in order to insure validity and reliability of the first examiner's interpretations. There was 90% coding agreement for Spanish samples and 93% coding agreement for English samples.

2.5.2 Sensitivity and specificity. According to Archibald (2009), sensitivity is defined as "the accuracy of a test in detecting individuals with the disorder" (pg 900). In context of the current study, the sensitivity equates to how accurate the BESOS was in determining whether impairment persisted (true positive). Specificity, on the other hand, was defined as "the accuracy in identifying those without the disorder" (pg 900) (Archibald, 2009). For this study, the specificity was the accuracy of the BESOS in saying that the participant was not at risk for LI, when LI truly did not persist (true negative). Sensitivity and specificity were calculated to determine diagnostic accuracy of the BESOS when compared to the gold standard. Recommended sensitivity and specificity interpretations are seen in the following hierarchy: 90=Good, 80-89=Fair/acceptable, <80 =unacceptable (Plante and Vance, 1994). However, such guidelines are suggested for assessments not screeners. Lugo-Neris et al. (2015) suggests using lower criteria for screeners, tolerating levels between 70% and 80%, as proposed by Barnes, 1982 as we expect some over-identification of these tools.

Chapter 3: Results

Information gathered from the PLS-5, LSA, and ITALK were used to determine if language impairment (LI) was present. First, the scores of each participant were examined to determine if they "passed" or "failed" the 3 tasks. The scores on the PLS-5, LSA, and questionnaires were then compared to BESOS to answer the research questions by calculating the sensitivity and specificity overall and for each age group.

3.1 Screening Result Measures

3.1.1 BESOS. For the Pre-School group, the language demonstrating strongest scores varied on the BESOS (see Table 3.1.1). Performance on the Semantic Subtest of the BESOS was split equally between languages; half of participants in this group performed higher in English (4/8), while half performed higher in Spanish (4/8). Performance on the Morphosyntax Subtest for this age group, however, demonstrated that participants overwhelmingly performed higher when assessed in English (7/8), than in Spanish (1/8). For the Pre-Kinder group, the majority of participants performed better on the English administration (5/6) than the Spanish administration (1/6) for the Semantic Subtest. All Pre-Kinder performed higher in English for the Morphosyntax Subtest (6/6). Similarly, all of the participants in the Kindergarten group scored higher in English (5/5) on the Semantics Subtest and the majority performed higher in English (4/5) in Morphosyntax.

Using a <-1 z-score as a "fail" or "at risk" criteria, a participant must have scored below the cut-off for one out of the two domains. Scores > -1 z-score resulted in a pass or "not at risk" (Lugo-Neris, 2015). Table 3.1.1 results depict a 1, representing a pass and a 0, representing a fail. Results indicated that 6/8 participants in the Pre-School group were "at risk" for LI, and

failed the screener. All participants in the Pre-Kinder and Kinder groups were not "at risk" for LI and passed the screening measure.

Table 3.1.1
BESOS Results

I.D.	*Semantics*		*Morphosyntax*		*P/F*
	English	*Spanish*	*English*	*Spanish*	
Pre-school					
1	**-.5**	-1.19	**1.3**	-2.52	1
2	**1.56**	-2.97	-1.92	**.51**	1
3	-2.97	**-1.64**	**-1.92**	-2.24	0
4	-2.97	**-2.56**	**-1.92**	-2.52	0
5	-2.97	**-1.19**	**-1**	-2.52	0
6	-2.27	**-1.65**	**-1.91**	-1.97	0
7	**-1.49**	-3.02	**-.38**	-2.52	0
19	**-1.37**	-3.48	**-.77**	-2.09	0
Pre-Kinder					
9	**0**	-1.19	**-.54**	-2.52	1
11	**.5**	-1.64	**1.07**	-2.52	1
12	**-.5**	-3.02	**-.76**	-2.52	1
13	0	**.18**	**1.07**	-1.97	1
17	**.5**	-1.2	**-.08**	-2.52	1
18	**.45**	-1.65	**1.07**	-1.97	1
Kinder					
8	**2.2**	-1.19	**1.76**	-2.1	1
10	**.07**	-3.02	**1.53**	-2.31	1
14	**1.4**	-2.11	**1.07**	-2.31	1
15	**.96**	.18	1.07	**1.49**	1
16	**1.31**	-2.37	**1.78**	-2.14	1

Note. Bold text= best language z-score for subtest; P/F= pass/fail; Pass=1, Fail=0; I.D.= participant identification number

3.2 Gold Standard Result Measures

3.2.1 PLS-5. Standard scores for the Auditory Comprehension and Expressive Communication subtests are reported in Table 3.2.1 for the English or Dual Language version of the PLS-5. One participant scored below -1 SD for the Expressive Communication subtest, however, was within 1 SD for the Total Language Score, indicating language skills were within the average range. All participants passed the PLS-5 based on the -1SD criteria and were shown to have typical language for their age and gender based on this assessment.

Table 3.2.1
PLS-5 Results

I.D.	*English Standard Score*		*Dual Standard Score*		*Total Language score*	*P/F*
	AC	*EC*	*AC*	*EC*		
Pre-school						
1	-	-	130	133	134	1
2	-	-	139	120	132	1
3	-	-	103	92	97	1
4	-	-	99	85	91	1
5	-	-	103	104	104	1
6	-	-	107	98	103	1
7	94	*84*	-	-	88	1
19	92	90	-	-	90	1
Pre-Kinder						
9	108	94	-	-	101	1
11	101	92	-	-	96	1
12	94	90	-	-	91	1
13	-	-	125	126	128	1
17	101	88	-	-	94	1
18	128	150	-	-	142	1
Kinder						
8	-	-	133	138	138	1
10	109	110	-	-	110	1
14	127	118	-	-	124	1
15	-	-	115	125	122	1
16	102	110	-	-	106	1

Note. AC= Auditory comprehension subtest; EC= Expressive Communication Subtest; P/F= pass/fail; Pass=1, Fail=0; I.D.= Participant identification number

3.2.2 LSA. All data was transcribed through SALT Software (Miller, Ansriacchi, & Nockerts, 2011). Results showed that 4 out of 8 children in the Pre-School group produced less than a cut-off of 58.38% grammatical utterances; 2 out of 6 in the Pre-Kinder group scored below the cut-off; and no participants scored below the cut-off in the Kinder group (see Table 3.2.2). The majority of participants completed the sample in English only (12/19); 5 out of 19 completed the sample in Spanish and two participants completed a language sample in both languages. Results showed that most individuals that scored below the cut-off were in the Pre-School group and were tested in Spanish, contributing to four of the six "failures" for this task.

Table 3.2.2
LSA Results

I.D.	*English Percent Grammatical Utterances*	*Spanish Percent Grammatical Utterances*	*P/F*
Pre-school			
1	69%	-	1
2	-	74%	1
3	-	41%	0
4	-	9%	0
5	-	25%	0
6	-	55%	0
7	84%	-	1
19	70%	-	1
Pre-Kinder			
9	56%	-	0
11	73%	-	1
12	44%	-	0
13	80%	-	1
17	82%	-	1
18	81%	-	1
Kinder			
8	89%	55%	1
10	80%	-	1
14	79%	-	1
15	70%	76%	1
16	94%	-	1

Note. P/F= pass/fail; Pass=1, Fail=0; I.D.= Participant identification number

3.2.3 ITALK. For English, 6 of 19 parents rated participants higher than teachers. However, the majority (10/19) of parents and teachers rated participants similarly (< .20 difference between teacher and parent ratings) for all age groups. The remaining three were rated higher by teachers. Spanish revealed a different pattern. Parents rated the participants higher than teachers did for Spanish proficiencies (10/19) in all age groups. Parents and teachers similarly rated 5 of the 19 participants. Participants "failed" this section of the screener if the participants' higher language was below a cut-off of 4.18. Of the participants whose scores fell below the cut-off, the majorities (3/5) were in Pre-School and were Spanish dominant. The other 2 participants who did not pass this task were English dominant, (1 Pre-Kindergarten and 1 Kindergarten). All participant proficiency ratings are reported in Table 3.2.3.

Table 3.2.3
ITALK Proficiency ratings

I.D.	*English Parent Rating*	*Spanish Parent Rating*	*English Teacher rating*	*Spanish Teacher rating*	*P/F*
Pre-school					
1	4.8	3.4	4.8	3.4	1
2	3.6	4.6	2	5	1
3	2	4.2	1.4	3.2	0
4	3.2	4	1.6	2.8	0
5	2.6	4.8	2.4	3.6	0
6	3.6	5	3.8	5	1
7	5	2.8	5	2.6	1
19	4.8	3.2	5	3	1
Pre-Kinder					
9	4	4.6	4.8	2	1
11	4.2	3.8	5	2.4	1
12	4.2	3.6	4.4	2.5	1
13	5	4.8	4.4	4.4	1
17	4.6	2.8	4	3.2	0
18	4.8	3.8	4.8	5	1
Kinder					
8	4.8	3.6	4.4	2.4	1
10	5	1.8	4.8	2.6	1
14	4.8	2.6	4	2.2	0
15	3.8	5	3.8	4.6	1
16	4.2	2.2	4.4	2	1

Note. Ratings are based on a 5-point Likert scale*;* P/F= pass/fail; Pass=1, Fail=0; I.D.= Participant identification number

3.3 Diagnostic Accuracy

To determine the diagnostic accuracy of the BESOS, scores from the BESOS were compared to the gold standard tasks (PLS-5, LSA, and ITALK). A passing score on the gold standard tasks meant that participants passed at least 2 out of the 3 tasks. Of the 19 participants examined, 6 participants were determined to be at risk for LI based on their BESOS scores. The gold standard, however, revealed that only 3 of those participants truly had LI. Revealing that 3 participants were misidentified by the BESOS. Results showing BESOS pass/fail scores along with an overview of pass/fail scores for each gold standard assessment are shown in Table 3.3.1.

Table 3.3.1
Pass/Fail Results Summary

I.D.	***PLS-5 P/F***	***LSA P/F***	***ITALK P/F***	***LI P/F***	***BESOS P/F***
Pre-school					
1	1	1	1	1	1
2	1	1	1	1	1
3	1	0	0	0	0
4	1	0	0	0	0
5	1	0	0	0	0
6	1	0	1	1	0
7	1	1	1	1	0
19	1	1	1	1	0
Pre-Kinder					
9	1		1	1	1
11	1	1	1	1	1
12	1	1	1	1	1
13	1	1	1	1	1
17	1	1	0	1	1
18	1	1	1	1	1
Kinder					
8	1	0	1	1	1
10	1	1	1	1	1
14	1	1	0	1	1
15	1	1	1	1	1
16	1	1	1	1	1

Note. P/F= pass/fail; Pass=1, Fail=0; I.D.= Participant identification number

3.4 Analysis

3.4.1 Overall sensitivity and specificity of BESOS. Overall, the BESOS had a specificity of 81% and sensitivity of 100%. The BESOS identified 13 participants as not being at risk for LI and not having LI by the gold standard (true negatives). No false negatives were found. Additionally, three participants were at risk for LI and did have LI (true positives). Three participants were identified as at risk for LI but passed the gold standard, resulting in false positives.

3.4.2 Sensitivity and specificity between age groups. A specificity of 40% and sensitivity of 100% were found for the Pre-School group (See Table 3.4.1). Two participants were not at risk for LI, and were not identified as having LI by the gold standard (true negatives). All true positives and false positives mentioned above were in the Pre-School group.

The BESOS results showed greatest specificity for the Pre-Kinder and Kinder groups, resulting in 100% specificity for each group. However, sensitivity could not be calculated for

these groups as no children had LI, as determined by the gold standard. All Pre-Kinder and Kinder group participants passed the BESOS.

Table 3.4.2
Sensitivity and Specificity of the BESOS

	Sensitivity	*Specificity*
Overall	*100%*	*81%*
Pre-School	*100%*	*40%*
Pre-Kinder	[a]	*100%*
Kinder	[a]	*100%*

Note. [a] sensitivity could not be determined.

Chapter 4: Discussion

In the past couple of decades, the importance of effective bilingual assessment and screening methods has received great attention in the literature. Children who are bilingual are often misidentified with LI when truly the characteristics seen in assessment may be differences in the language itself or a language-learning gap (Paradis, Schneider, & Duncan, 2013). The recognition of the mismatch seen in practice has led to newer developments of screening practices and strategies in assessment tools with bilingual considerations (Patterson et al. 2013; Sheng et al. 2012; Peña et al., in preparation). The purpose of this pilot study was to explore the accuracy of the BESOS for use in the El Paso, Texas region.

4.1 Validity findings

4.1.1 Overall validity of the BESOS. Following recommendations made by Plante and Vance (1994), the BESOS' overall sensitivity (100%) was "good" and specificity (81%) was "fair/acceptable". Although specificity was on the lower range according to Plante and Vance (1994), this is not uncommon for a screener. By design, screeners aim to over-identify children at risk for LI (Barnes, 1982). Results from the current study in El Paso are consistent with a predictive sensitivity of 95.2% and a predictive specificity of 71.4% for the BESOS norming sample (Lugo-Neris et al., 2015).

Compared to other pre-school language screeners, the results from the BESOS in the current study had a different pattern of sensitivity and specificity. The current study had a higher sensitivity than specificity. In other words, the BESOS was more accurate in identifying individuals *without* LI, than those *with* LI. Literature investigating other available language screeners found a higher specificity than sensitivity. For example, the Preschool Language Scales Screening Test-5th Edition has a sensitivity of 86% and specificity of 96% (Zimmerman, Steiner,

& Pond, 2012). Additionally the Sentence Repetition Screening Test (Sturner, Funk & Green, 1996) has a sensitivity of 76% and specificity of 92% and the Test for Examining Expressive Morphology (Persona, Plante & Vance, 2005) has a sensitivity of 90% and specificity of 100%. The BESOS shows overall promise in its validity as a test, when compared to standards and other screening measurements.

4.1.2 Validity of BESOS between groups. The BESOS had "good" sensitivity for all groups (Pre-School, Pre-Kinder, Kinder). Sensitivity was consistently 100% at accurately identifying participants without LI across all groups. However, because there were no participants with LI in the Pre-Kinder and Kinder groups, specificity could not be determined. Therefore, specificity results are inconclusive for these two groups. In the Pre-School group, on the other hand, the BESOS had a specificity of 40%. In other words, the BESOS accurately identified 40% of participants as "at-risk" when the participants truly had LI. These findings, however, are not surprising for the Pre-School group given the age of the participants. Although the BESOS contains some 3-year-olds in its norming sample, it is primarily developed to screen 4- to 6-year-old children. For the Pre-School group in the current study, ages ranged from 3;2-3;11 (see Table 2.1.1 for mean ages).

4.2. Future Research

Additional empirical research is needed in the use of preschool screening instruments such as the BESOS for bilingual children. The current study is a pilot study and will continue with additional participants in the future. It is suggested that other instruments undergo the same types of scrutiny as the one described above in order to find an adequate tool to use for English-Spanish bilinguals. Test developers are also encouraged to take a second language into consideration to make language screening and assessment as effective as possible. Future

research should include larger sample sizes, multiple testing locations, and participants with LI for all age groups.

4.3 Threats and Limitations

One of the greatest limitations of the current study was a small sample size. With nineteen participants, data could not be generalized to the entire El Paso population. However, this was exploratory by design, so results are preliminary. Yet, the data collection and analysis of this pilot study has already informed the larger study that is ongoing. Additionally, there were an unbalanced number of participants in each of the task administration sequences when randomly assigned to a condition and all data collection was gathered at only one location.

Due to data collection at a single location, possible extraneous factors could have impacted the results. Socio-economic status (SES) for the region could potentially play a role on outcomes, as well as language exposure. The facility was a dual language program, in which teachers instructed the children in English half of the time and in Spanish the other half. Results may look different at other facilities where Pre-School or Pre-Kinder children are not getting a formal education at their age or if the kindergarten children are assessed in a program where they were only exposed to a single language. It is important to note that as the participants received a monetary reward for their participation, this study may hold a bias as a result of having received an incentive as it may have motivated certain populations to participate versus others. For example, parents of children who come from a low SES may have been more enticed to return consent forms to participate.

Along with small sample size and location, there were few kids with language impairment in the sample. In the Pre-Kinder and Kinder groups none of the children had language impairment, making sensitivity inconclusive, as it could not be calculated for those

ages. Not having children with language impairment additionally was a limitation to the study as there was weaker evidence to support the specificity of the BESOS for the region.

4.4 Other Considerations

A major consideration that arose in this study was the accuracy of the gold standard. When assessing participants in all groups. The PLS-5 did not identify language concerns for any participants while the LSA and ITALK measures did. The PLS-5 findings support the use of three measures (PLS-5, LSA, ITALK) to determining if LI was present. Participants appeared to have no difficulty with PLS-5 stimuli. However, when in spontaneous conversation and other assessment measures, observable concerns were apparent to clinicians.

In addition to the PL5-5, some factors were additionally recognized for the LSA assessment. When gathering language samples, some participants had observably more difficulty than others. Thus, samples of longer length may be more representative of the participant's language than shorter samples. Possibilities for varying lengths in transcripts may have been due to abstract picture stimuli or shyness. However, it is important to recognize the lack of uniformity in sample length when determining if LI was present.

Lastly, considerations to take into account are patterns found in ITALK results. Overwhelmingly, parents rated children higher than teachers for Spanish. In English, however, ratings between parents and teachers were consistent with one another. Factors to consider are whether the children were truly exposed equally to both English and Spanish at school or whether parents may hold biases in the dominant/ non-dominant language. Other possibilities include parents' and possibly teachers' habituation to the child's language and lack of knowledge of typical language development, thus not recognizing concerns.

4.5 Conclusion

Results are contradictory in supporting the author's hypotheses. The BESOS screening test was accurate in identifying bilingual preschool children at risk for language impairment. However, unlike the author's hypothesis, there were differences in accuracy between specific age groups. The results support the need for further research in the use of language screeners for English-Spanish bilinguals. Due to the lack of specificity for two of the age groups, it is suggested that further screening continue in validating the BESOS. Results additionally show that although findings are inconclusive at the present time, the BESOS holds promise as an effective indicator in identifying Pre-School children who are at risk for language impairment in the El Paso, TX region. Shown by high sensitivity and an overall specificity of 81%, the BESOS shows most promise for 4 and 5 year old bilingual children in the El Paso, TX region. However, clinicians should be cautious in the populations as preliminary findings demonstrate less promise for 3-year-old populations. Based on the findings of this study, the accuracy of the BESOS is not to the standards they should be for testing 3 year olds.

References

Allen, D. V., & Bliss, L. S. (1987). Concurrent validity of two language screening tests. *Journal of Communication Disorders*, *20*(4), 305-317.

Archibald, L. M., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 899-914.

Beardsmore, H. B. (1986). *Bilingualism: basic principles* (Vol. 1). Multilingual Matters.

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

Bedore, L. M., Peña, E. D., Garcia, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference?. *Language, Speech, and Hearing Services in Schools*, *36*(3), 188-200.

Bedore, L. M., Pena, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K. A. I., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children.*Bilingualism: Language and Cognition*, *15*(03), 616-629.

Bernal, E. M., & Tucker, J. A. (1981). *A Manual for Screening and Assessing Students of Limited English Proficiency*.

Brebner, C., Chandler Yeo, H., Goh, M. M., Kam, K. W., & Yeo, W. S. F. (2015). Tests over time: Evaluating the currency of normative data in a complex multilingual environment. *International Journal of Speech Language Pathology*, *17*(6), 556-564.

Camilleri, B., Hasson, N., & Dodd, B. (2014). Dynamic assessment of bilingual children's language at the point of referral. *Educational & Child Psychology*, *31*(2), 57-72.

Curtis, C., Summers C., Smith, V., Stubbemann, R., (In Progress)

Early Childhood Intervention Services. (n.d.). Retrieved May 12, 2016, from http://www.dars.state.tx.us/ecis/

Eisenberg, S. L. & Guo, L. (2015). Sample size for measuring grammaticality in preschool children from picture-elicited language samples. *Language Speech Hearing Services in Schools*, *46*(2), 81-93. doi: 10.1044/2015_LSHSS-14-0049.

Eriksson, M., Westerlund, M., & Miniscalco, C. (2010). Problems and limitations in studies on screening for language delay. *Research in developmental disabilities*, *31*(5), 943-950.

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *54*(3), 280-290.

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological assessment*, *6*(4), 304.

Grech, H., & Dodd, B. (2011). Assessment of speech and language skills in bilingual children: an holistic approach. *Stem-, Spraaken Taalpathologie*,*15*(2).

Gross, M., Buac, M., & Kaushanskaya, M. (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American journal of speech-language pathology*, *23*(4), 574-586.

Guiberson, M., & Rodríguez, B. L. (2010). Measurement properties and classification accuracy of two Spanish parent surveys of language development for preschool-age children. *American Journal of Speech-Language Pathology*, *19*(3), 225-237.

Guiberson,M.,& Rodgriguez,B.(2014). Rasch analysis of a Spanish language screening parent survey. *Research in Developmental Disabilities, 35*, 646-656.

Gutierrez-Clellen, V. F., & Pena, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech, and Hearing Services in Schools*, *32*(4), 212-224.

Herdman, M., Fox-Rushby, J., & Badia, X. (1997). 'Equivalence'and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*, *6*(3).

Hoff, E., & Tian, C. (2005). Socioeconomic status and cultural influences on language. *Journal of Communication Disorders*, *38*(4), 271-278.

Katz, L. A., Maag, A., Fallon, K. A., Blenkarn, K., & Smith, M. K. (2010). What makes a caseload (un) manageable? School-based speech-language pathologists speak. *Language, Speech, and Hearing Services in Schools*,*41*(2), 139-151.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the literature. *Developmental Medicine & Child Neurology*, *42*(03), 190-200

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (1997). Screening for speech and language delay: a systematic review of the literature. *Health technology assessment (Winchester, England)*, *2*(9), 1-184.

Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., & Gillam, R. B. (2015). Utility of a language screening measure for predicting risk for language impairment in bilinguals. *American Journal of Speech-Language Pathology*, *24*(3), 426-437.

Miller, R., & Acton, C. (2009). *SPSS for social scientists*. Palgrave Macmillan.

Miller, J.F., Andriacchi, k., &Nockerts, S. (2011). *Assessing language production using SALT software: a clinician's guide to language sample analysis*. Middleton, WI:SALT Software LLC.

Norris, M. K., Juarez, M. J., & Perkins, M. N. (1989). Adaptation of a screening test for bilingual and bidialectal populations. *Language, Speech, and Hearing Services in Schools*, *20*(4), 381-390.

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment among English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, *56*(3), 971-981.

Patterson, J. L., Rodríguez, B. L., & Dale, P. S. (2013). Response to dynamic language tasks among typically developing Latino preschool children with bilingual experience. *American Journal of Speech-Language Pathology*, *22*(1), 103-112.

Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language learning*, *43*(1), 93-120.

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2014). *Bilingual English-Spanish assessment.* San Rafael, CA: AR-Clinical Publications.

Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development*, *78*(4), 1255-1264.

Perona, K., Plante, E., & Vance, R. (2005). Diagnostic Accuracy of the Structured Photographic Expressive Language TestThird Edition (SPELT-3). *Language, Speech, and Hearing Services in Schools*, *36*(2), 103-115.

Prelock, P. A., Hutchins, T., & Glascoe, F. P. (2008). Speech-language impairment: how to identify the most common and least diagnosed disability of childhood. *The Medscape Journal of Medicine*, *10*(6), 136.

Saville, D. J., & Wood, G. R. (1991). Randomized block design in statistical methods: The geometric approach. *Springer New York*. 299-339.

Sheng, L., Peña, E. D., Bedore, L. M., & Fiestas, C. E. (2012). Semantic deficits in Spanish–English bilingual children with language impairment. *Journal of Speech, Language, and Hearing Research*, *55*(1), 1-15.

Shipley, K.G., MacAfee, J.G. (2004) Assessment in speech language pathology. A resource manual third edition. *Delmar Cengage Learning*.

Sturner, R. A., Funk, S. G., & Green, J. A. (1996). Preschool speech and language screening: further validation of the sentence repetition screening test. *Journal of Developmental & Behavioral Pediatrics*, *17*(6), 405-413.

Tomblin, J. B., Smith, E., & Zhang, X. (1997). Epidemiology of specific language impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders*, *30*(4), 325-344.

U.S. Census Bureau. (2014). State & county Quickfacts: El Paso County, TX. Retrieved October 30, 2015, from http://quickfacts.census.gov.

Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2012). Pre- school Language Scales Screening Test–Fifth Edition. Bloomington, MN: NCS Pearson.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2012). Pre- School Language Scales–Fifth Edition. Bloomington, MN: NCS Pearson.

Curriculum Vita

Roxana Stubbemann was born in Sulingen, Germany and was raised in El Paso, Texas, USA. She is the second child and daughter of Mr. Juergen Stubbemann and Mrs. Theresa Stubbemann. She is a three-year high school graduate and without a Bachelors degree, completed pre-speech language pathology undergraduate coursework at the University of Texas at El Paso (UTEP). In the fall semester of 2015, she entered the UTEP Speech-Language Pathology program, to obtain her Master of Science degree. Roxana has been awarded numerous awards within her time in the graduate program including, a summer research assistantship, Dodson Research Grant, and Jimmie Vokes Bernard Endowed Scholarship. As a member of the Research in Bilingual Language Learning Lab, she has presented her research at the UTEP Graduate Student Research Expo. She is additionally a student member of the Texas Speech Hearing Association, El Paso Speech Hearing Association, and National Student Speech Language Hearing Association, in which she is currently a communication officer.

Permanent Address: 11365 East Ranch Ct.
El Paso, Tx 79934
rstubbemann@miners.utep.edu

This thesis was typed by Roxana Stubbemann.