10-8-2017

A Comparison of Pass Rates Across Three Language Screeners for Spanish-English Bilingual Children

Mariajose Bosanko
Louisiana State University and Agricultural and Mechanical College, mbosan1@lsu.edu

A COMPARISON OF PASS RATES ACROSS THREE LANGUAGE SCREENERS FOR SPANISH-ENGLISH BILINGUAL CHILDREN

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
Requirements for the degree of
Masters of Arts

in

The Department of Communication Sciences and Disorders

by
Mariajosé Bosanko
B.F.A., Kendall College of Art and Design, 2002
M.A., University of New Mexico, 2008
December 2017

TABLE OF CONTENTS

ABSTRACT

There is limited research regarding the efficacy of the language screening process, especially for bilingual populations. The purpose of this study was to examine the pass rates of three language screeners when administered to Spanish-English bilingual and ELL children. A total of ten Spanish-English bilingual children enrolled in either Pre-K, kindergarten, or first grade completed each screener. The screeners were: the *Preschool Language Scales Spanish Screening Test-Fifth Edition* (PLSSST-5; Zimmerman et al., 2012a), the *Bilingual English Spanish Oral Screener* (BESOS; currently in development; Lugo-Neris et al., n.d.), and the *Diagnostic Evaluation of Language Variation-Screening Test* (DELV-S; Seymour et al., 2003). Analyses included examining each tool's pass rate, the consistency at which the tools identified the same participants as requiring further evaluation, and the accuracy at which the tools classified the children's referral status (+referral vs. –referral) as determined by the *Instrument to Assess Language Knowledge* (ITALK; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014). Thirty percent of the participants passed the PLSSST-5 and the DELV-S-II, and 60% passed the BESOS. When compared to the ITALK, the PLSSST-5 yielded the highest classification accuracy (90%) for identifying the participants' referral status. The results suggest caution when using these language screeners by themselves to determine if a child should be referred to a speech-language pathologist for evaluation.

CHAPTER 1
INTRODUCTION

As the Spanish-speaking population in the United States continues to grow, accurately identifying whether Spanish-English bilingual children should be referred for a language evaluation is an important issue. According to the U.S. Census Bureau (2014), 21% of the population 5 years of age and over speaks a language other than English, with Spanish (13%) being the largest language group. To appropriately identify children who need a comprehensive language evaluation, it is important that language screening instruments are available.

Language screeners are supposed to be brief and effective tools to help identify children who should be referred for further testing. As a screener, administration should require minimal effort for both the administrator and child (Lugo-Neris et al., 2015). Screening for a language impairment can take various formats, including direct assessments through norm-referenced or criterion-referenced instruments, informal testing, observation, caregiver and/or teacher report, developmental checklists, or some aggregate of these.

Hosp and Reschly (2003) contend that referral for further assessment is thought to be the most significant step in the process of determining a child's eligibility for intervention services, while also stating that there are shortcomings in the reliability and validity of the screening process. Use of standardized screeners as compared to informal ones should help provide more reliability and validity to the referral process. Additionally, screeners that are designed specifically for linguistically diverse populations should reduce cultural biases that may be inherent to instruments designed for other populations.

Currently, there are two language screening instruments whose norming samples include Spanish-English bilinguals and English language learners (ELLs): the *Preschool Language Scales Spanish Screening Test-Fifth Edition* (PLSSST-5; Zimmerman et al., 2012a) and the

Bilingual English Spanish Oral Screener (BESOS; currently in development; Lugo-Neris et al., n.d.). The *Diagnostic Evaluation of Language Variation-Screening Test* (DELV-S; Seymour et al., 2003) is a third instrument that may be appropriate for Spanish-English bilingual children because it includes a small sample of speakers of Spanish-influenced English. Both the PLSSST-5 and BESOS are norm-referenced, and the DELV-S is criterion-referenced.

The current study was designed to learn more about the concurrent validity of the PLSSST-5, BESOS, and DELV-S when these tools are administered to children who are bilingual and/or ELL. To do this, I examined each tool's pass rate, the consistency at which the tools identified the same participants as requiring further evaluation, and the accuracy at which the tools classified the children's referral status (+referral vs. –referral) as determined by the *Instrument to Assess Language Knowledge* (ITALK; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014). The ITALK is a caregiver and teacher rating tool that has been designed for bilinguals and ELLs. As background, I review the terminology and labels in studies of children who are bilingual, details about the development of the three language screening instruments examined, and previous studies of existing language screeners. As evident from the review, although screening for language impairment is an important first step in the identification of those who require a full evaluation, few studies have examined the concurrent validity of screening tools.

CHAPTER 2
LITERATURE REVIEW

Varying Linguistic Profiles of Children Who Are Bilingual

The screening instruments used for this study have different norming samples that reflect the diversity of children who are bilingual and/or ELLs. Screeners may be selected for use based on language proficiency determined by classification categories or labels. Classification practices and labels differ significantly; because of this it is important to review the labels and terminology used in the literature.

Bilinguals. Valdés and Figueroa (1994) describe bilinguals as those who have knowledge of more than one language. Bilinguals can be categorized as simultaneous or sequential. Simultaneous bilinguals acquire two or more languages at the same time before age three, and sequential bilinguals acquire one language prior to another (Paradis, Genesee, & Crago, 2011). Students who first come into contact with a second language at school are considered sequential bilinguals. Some researchers also refer to this group as ELLs or emergent bilinguals (García & Kleifgen, 2010).

In an attempt to further categorize the bilingual and ELL population, other labels are sometimes used to describe children's expressive and receptive capabilities (e.g., expressive bilingual, receptive bilingual) or their comparative abilities of their two languages (e.g., Spanish dominant, English dominant, balanced bilingual; Quin Yow & Li, 2015). Attaching these more specific labels to children poses a challenge in research and clinical practice because children's bilingual language experiences are so varied. Bedore et al. (2012) further note that bilingual children exhibit ability differences across domains of language, and their abilities can shift across time. Given this, it is possible for someone to move from one label to another. Paradise et al.

(2011) also state that while these labels help to make distinctions among bilingual/ELL research participants, these distinctions are not precise or mutually exclusive.

Speakers of Spanish-influenced English. Not all speakers of Latino[1] ethnicity speak Spanish. Speakers of Spanish-influenced English may have Spanish as their heritage language, but they may not speak Spanish. Instead, the English they have learned is a variety that has been influenced by Spanish (Gutiérez-Clellen et al., 2008; Santa Ana A., 1993). Spanish-influenced English presents characteristics of Spanish in morpho-syntax, phonology, and lexical attributes, and it shares some qualities with other English dialects, such as African American English (AAE; Gutiérez-Clellen et al., 2008). One variety of Spanish-influenced English is Chicano English. Chicano English specifically refers to an influence on English that is tied to Mexican Spanish. Children may also speak other dialects of Spanish-influenced English if the Spanish variety reflects a variant that differs from Mexican Spanish (e.g., Puerto Rican English and Cuban English), but much of what is known about Spanish-influenced English has been based on Chicano English.

Identifying Spanish-influenced English as a dialect is a matter of contention. Some argue that it is not a dialect but rather a state of incomplete acquisition of English. Others have shown that Spanish-influenced English adheres to dialect-specific rules that differ from those of a Spanish-English bilingual system (Gutiérrez-Clellen et al., 2007; Santa Ana A., 1993). From a sociolinguistic perspective, language contact situations in which bilingualism occurs often

[1] The terms Hispanic and Latino are commonly viewed as interchangeable, with the term Hispanic being used by Governmental departments in the US. However, the two can have different meanings by the people who are given these labels. The author prefers the term Latino, derived from Latin America, because it denotes diversity in national origin and race, and a similar treatment of this population as a group in the US (Hayes-Bautista and Chapa, 1987; Amaro and Zambrana, 2000). The term Hispanic denotes a Spanish or Iberian cultural origin regardless of race (Martin-Alcoff, 2005).

change across time, especially when one language is more dominant than the other in a community. In these situations, as more speakers begin to use the dominant language and lose their ability to speak their heritage language, the dominant language begins to evolve and change. Dubois and Horvath (2003) have documented this type of language evolution for communities of Cajun English speakers. For these reasons, Spanish-influenced English like Cajun English and AAE, can be argued to reflect a dialect of English rather than a transitional variety.

Given the heterogeneity that exists within bilingual communities, the current study included caregiver and teacher interviews to describe the language status of the participants. These interviews were guided by the *Bilingual Input-Output Survey* (BIOS; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014). The BIOS allows a researcher to quantify a child's use of Spanish and English in school and home.

Three Language Screeners

PLSSST-5. The PLSSST-5 is the only commercial norm-referenced Spanish-English bilingual language screener for children between the ages of 3;0 to 7;11. The PLSSST-5 consists of six speech and language subtests: articulation, language, connected speech, social/intrapersonal communication skills, fluency and voice. These items were derived from the *Preschool Language Scales-5* (Zimmerman et al., 2012b) and the *Preschool Language Scales-5 Spanish* (PLS-5 Spanish, Zimmerman et al., 2012c). The PLSSST-5's normative sample of 1150 children included two types of language users; 67% were Spanish speakers and 33% were Spanish-English bilingual speakers who were considered Spanish dominant (Zimmerman et al., 2012c).

BESOS. The BESOS is a screening instrument that is currently under development. It consists of four subtests: English and Spanish morpho-syntax and English and Spanish semantics. The subtests were derived from the *Bilingual English Spanish Assessment* (BESA, Bedore et al., 2012), an instrument developed to identify bilingual 4- to 6-year-olds with speech and language impairments.

The BESOS was normed on 800 Spanish-English bilingual children growing up in the US who varied in their experiences and use of each language. The normative sample for each language subtest included children reported to use that language at least 40% of the time on a daily basis.

DELV-S. The DELV-S is a language screener designed for children who speak a variety of English dialects. It consists of two subtests: language variation status (DELV-S-I) and diagnostic risk (DELV-S-II; Seymour, deVilliers, & deVilliers, 2003). The DELV-S-I classifies the dialect of the child as either mainstream American English (MAE), some variation from MAE, or strong variation from MAE. The DELV-S-I is for ages 4;0 to 12:11, and the DELV-S-II is for ages 4;0 to 9;11. The DELV-S-II estimates a child's risk for a language impairment using the categories: lowest risk for language impairment, low to medium risk, medium to high risk, and high risk. The DELV-S' normative sample included 1,258 children from different regions of the US, and it included an over-sampling of speakers of AAE and a small sample of speakers of other English varieties, including Spanish-influenced English.

Table 2.1 compares and contrasts the three instruments. As summarized in the table, all three screening instruments have a morpho-syntactic component. The PLSSST-5 has these components as part of the language subtest, the BESOS has a morpho-syntax subtest for each language, and the DELV-S targets this area of language in both subtests. The PLSSST-5 includes

an articulation subtest that consists of 41 target phonemes. The BESOS and DELV-S-II do not assess children's articulation abilities. Additionally, the DELV-S-II has non-word items that children are asked to repeat. These items consist of nonsense words that were created to follow the phonotactic rules of any English variety (Seymour, deVilliers, & deVilliers, 2005).

Table 2.1 Screening instrument comparison

	PLSSST-5	BESOS	DELV-S-II
Subtests	language, articulation, connected speech, social/intrapersonal communication skills, fluency and voice	Spanish morpho-syntax, English morpho-syntax, Spanish semantics, English semantics	WH-questions, noncontrastive grammar structures, repeat nonwords
Parent test	PLS-5 and PLS-5 Spanish	BESA	DELV-NR
Language use	Scripted direct translation of Spanish and English	Not direct translation of Spanish and English	English only
Examiner qualifications	Native or near native Spanish fluency, Spanish-English bilingual, qualified interpreter	Bilingual examiner proficient in both Spanish and English	Professionals with experience in assessment
Administration	First administered in Spanish and subsequently incorrect items administered in English	Administer each language subtest individually, allowing for responses in either language on semantics subtest	Both subtests are administered
Scoring	Scores can be derived from Spanish only or dual language administration	Four subtests scores derived that can be combined or only use a combination of best scores	Lowest risk, low to medium risk, medium to high risk, and high risk

(Table continued)

	PLSSST-5	BESOS	DELV-S-II
Norming sample	770 monolingual Spanish speakers and 380 Spanish-English bilinguals who are Spanish dominant	800 bilingual children growing up in USA with a broad range of exposure to Spanish and English	1,258 children speakers of MAE and AAE. 80 children speakers of other varieties of English: Appalachian English, Cajun English, Southern English, and Spanish-influenced English
Sensitivity	95%	four scores 86%, best two scores 90%	Ages 5;0-5;11 73%, Ages 6;0-6;11 70%
Specificity	79%	four scores 71%, best two scores 71%	Ages 5;0-5;11 82%, Ages 6;0-6;11 77%

The PLSSST-5 and BESOS include semantic items. The BESOS has a designated subtest for semantics while the PLSSST-5 includes semantic items within the language subtest. The DELV-S-II does not assess children's semantic systems. The PLSSST-5 includes connected speech, social/intrapersonal communication skills, fluency and voice subtests, which are not included in either the BESOS or the DELV-S-II. For the purpose of this thesis, only the language section of the PLSSST-5 was examined, because this section most closely aligned with the content assessed on the other two screeners. Also, for the DELV-S, only the risk subtest (i.e., DELV-S-II) was examined. The language variation subtest (i.e., DELV-S-I) of this screener was administered to the children but it was used to describe the participants' English dialect.

The final information presented in the table relates to the three screeners' sensitivity and specificity. Sensitivity denotes the proportion of people who score at risk for language impairment and who actually are at risk, and specificity refers to the proportion who are not at risk and score as such. Since the PLSSST-5 and the DELV-S-II are commercially available, information about sensitivity and specificity are included in their respective manuals. By

comparison, information about sensitivity and specificity for the BESOS is preliminary and comes from an experimental study (Lugo-Neris et al., 2015). For diagnostic assessment tools, Plante and Vance (1994) recommend sensitivity and specificity values that are no lower than 90%. However, Lugo-Neris et al. (2015) contend that because screening instruments do not automatically result in a diagnosis, there can be more leniency in the standards and argue that specificity, but not sensitivity, can range from 70-80%. From a meta-analysis of school-based curriculum-based screeners, Kilgus et al. (2014) also recommend sensitivity values greater than or equal to 80% and specificity greater than or equal to 70% to ensure that children who need to be referred are not missed.

The PLSSST-5 has a reported sensitivity of 95% and specificity of 79%. The BESOS can be scored two different ways so it has different sensitivity and specificity values based on the scoring method. For a combination of all four scores, sensitivity is reported as 86%, and specificity is reported as 71%. When using the best score of each subtest, sensitivity is 90%, and specificity is 71%. The DELV-S-II reports sensitivity ranges of 52%-73%, and specificity ranges of 76%-90%. Ranges are provided in the test manual because values vary as a function of the children's ages. For the 5;0-5;11 year olds, sensitivity is reported as 73% and specificity is reported as 82%. For the 6;0-6;11 year olds, sensitivity is reported as 70% and specificity is reported as 77%. The sensitivity and specificity values of the three screening instruments are within or close to the Lugo-Neris et al. (2015) and Kilgus et al. (2014) ranges of acceptability.

ITALK as a Reference Standard

The Inventory to Assess Language Knowledge (ITALK; Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014) is a caregiver and teacher interview designed to determine the childrens' speech and language ability in Spanish and English. The ITALK consists of five

areas: Vocabulary, Speech, Sentence Production, Grammar, and Comprehension. Caregivers and teachers are asked to rate the child's level of performance in each language, and the results are quantified. The ITALK was selected as the reference standard following other studies (Gutiérrez-Clellen & Kreiter, 2003; Restrepo, 1998; Simon-Cerejido & Gutiérrez-Clellen, 2007). Restreop (1998) and Simon-Cerejido and Gutiérrez-Clellen (2007) also argue that due to a lack of standardized measures for bilingual populations, parent and teacher reports are an important tools for identifying childhood language impairments. These authors also found parent and teacher reports to yield moderate to high correlations with other measures (correlations ranged from .73 to .79).

Previous Studies of Language Screening Instruments

There are at least five studies that have compared language screeners to each other or to other criterion–based or norm-referenced tools. All five studies included children between the ages of 4;0 and 6;4 years. Only two studies specifically stated that the participants were from culturally and linguistically diverse backgrounds, and only one included bilingual children. Measures, methodology, and findings for each study were reviewed. As evident from the review, the results from these five screening studies have been less than ideal, at least when the screeners are administered and scored according to their manuals.

Blaxley et al.'s (1983) study included 90 children, aged 4;00-6;11. Their screeners were the *Fluharty Preschool Speech and Language Screening Test* (FPSLST; Fluharty, 1978) and the *Bankson Language Screening Test* (BLST; Bankson, 1990). These screeners were administered in random order and compared to the children's *Developmental Sentence Score* (DSS; Lee & Canter, 1971), which was used as the reference standard for determining the children's referral status (i.e., language impaired and should be referred vs. typically developing and should not be

referred). To calculate a DSS, the authors collected language samples with 50 sentences each, and these were scored using a cut score of the 10th percentile. By comparing the screening outcomes to the child's DSS scores, the authors determined false negative and false positive rates. False negatives referred to those who passed the screening but were in need of further assessment as indicated by their DSS, and false positives referred to those identified as needing further assessment but were typically developing as indicated by their DSS.

The authors found that the false negative and positive rates were 16% and 3% for the FPSLST, and 7% and 9% for the BLST. The false negative rates indicate that some children with language disorders would likely not be referred for further assessment. The two screeners differed in their pass rates and they did not always pass the same children, meaning that some children passed one screener and failed another.

Illerburn et al.'s (1985) study included 136 children, aged 4;8-6;4 years. In this study, the authors administered screeners and three months later they administered three different language assessment tools as their reference standard for determining the children's referral status. The screeners were: BLST, FPSLST*, Language Identification Screening Test for Kindergarten* (LIST-K; Illerbrun et al., 1984), *Kindergarten Language Screening Test* (KLST; Gauthier & Madison, 1998), and the *Clinical Evaluation of Language Functions-Elementary Screening Test* (CELF; Semel & Wiig, 1980). The diagnostic assessment tools were: the *Carrow Elicited Language Inventory* (CELI; Carrow, 1974), *Test of Language Development* (TOLD; Newcomer & Hammil, 1982) and *Test for Auditory Comprehension of Language* (TACL; Carrow-Woolfolk, 1985). They found that the LIST-K had the highest correlation with the diagnostic assessment tools. Individually, the percentage of children correctly classified by the screeners was: KLST 84%, BLST 94%, CELF 92%, FPSLST 90%, and LIST-K 92%. Overall, the misclassified cases

ranged from 6%-16% depending on the test. This number of misclassified cases were lower than the 6-19% of misclassified cases identified by Blaxely et al. (1983). However, the authors took all the incorrect classifications and divided them equally among false negatives and false positives assuming a random distribution. Additionally, the authors did not examine the consistency at which the five screeners passed or failed the children.

Moland's (2011) study consisted of 73 children, aged 4;00 to 5;25 years. All were African American and enrolled in Head Start or a publicly-funded preschool. The screeners were: the *Fluharty Preschool Speech and Language Screening Test-Second Edition* (Fluharty-2*;* Fluharty, 2001), DELV-S-II, and the *Washington-Craig Language Screener* (WCLS; Washington & Craig, 2004). Moland (2011) does not provide information about the screeners' false negative or false positive rates, because a diagnostic instrument was not used to determine the referral status of the participants. However, all were deemed typically developing by their schools. Instead, she examined the pass rates of each screener and the consistency at which the screeners passed or failed the same children. In total, 25% of the children passed all three screeners and 29% failed all three screeners. The remaining 46% of children inconsistently passed or failed the screeners. In this study, a greater percentage of participants failed the 1st and 2nd screener than the 3rd screener, regardless of which screener was given last. Moland (2011) interpreted this finding as showing a practice effect and recommended further evaluation for the children who failed the last screener administered.

Rodriguez and Guiberson's (2011) study included 353 preschool children, with an average age of 4;2 years. The language groups consisted of 34 Spanish-speakers, 109 bilingual speakers, and 210 English-speakers. A total of 16 teachers also participated; ten were bilingual Spanish-English speakers and six were monolingual English speakers who had bilingual teaching

assistants. The *Teacher Rating of Oral Language and Literacy* (TROLL; Dickinson et al., 2001) was the screener and the *Preschool Language Scale-4* (PLS-4; Zimmerman et al., 2002) was the norm-referenced diagnostic tool. The TROLL and PLS-4 were administered six weeks after the beginning of the school year.

For all children combined, the TROLL correlated with the PLS-4. For the English-speaking children, the TROLL also correlated with the PLS-4 but for Spanish-speaking children, only the receptive subscale of the PLS-4 correlated to the TROLL. For the bilingual children, there was no significant correlation between the TROLL and the PLS-4 scores. The authors also found that the TROLL scores for the Spanish-speaking and bilingual population led to over-referral of these groups, even when the teachers were bilingual themselves. However, the authors did not report information on false positive or false negative rates of the TROLL. Given this, the percentage of children who were incorrectly classified as +/-refer is unknown.

Finally, Gregory and Oetting (under review) compared the results for four screeners. Participants were eight teachers and 98 children whose average age was 5;4 years. Of the participating children, 56% were typically developing and 44% were language impaired as determined by the syntax subtest of the *Diagnostic Evaluation of Language Variation-Norm Referenced* (DELV-NR; Seymour et al., 2005). The screeners were: the DELV-S-II, TROLL, the *Dynamic Indicators of Basic Early Literacy Skills Next* (DIBLES; Good, Gruba, & Kamanski, 2009), and the *Children's Communication Checklist-2* (CCC-2; Bishop, 2006). Only 49% of the participants passed all the screeners and 3% failed all the screeners. The remaining 48% inconsistently passed or failed the screeners. When using the clinical cutoffs recommended in the screening manuals, the four tools' sensitivity ranged from 9% to 56%, and specificity ranged

from 75% to 98%. These results indicate that the accuracy of all four screeners was too low to be clinically useful.

In summary, there are a number of bilingual children in the US, and the number of bilingual children continues to grow. Speech-language clinicians are in need of valid language screeners to determine which bilingual children should be referred for a full evaluation. Currently, three language screeners exist for bilingual Spanish/English children. These include the PLSSST-5 and BESOS, which include Spanish and English subtests, and the DELV-S-II, which includes an English-only subtest that is advertised as appropriate for children who speak Spanish influenced-English. Finally, minimal research has been conducted on language screening instruments. None of the studies reviewed identified an ideal screener for use by speech-language pathologists, and only one of the studies included children who were bilingual.

The purpose of the current study was to learn more about language screening instruments by examining the validity of the PLSSST-5, BESOS, and DELV-S-II when given to a diverse group of bilingual and ELL children. To do this, I examined the pass rates of each screener and the consistency at which the screeners passed or failed the same children. Then, using the ITALK as the reference standard to determine the children's referral status (+referral or – referral), I examined the classification accuracy of the screeners. The following questions guided the study:

1. What proportion of participants pass the PLSSST-5, BESOS, and DELV-S-II?
2. Do the PLSSST-5, BESOS, and DELV-S identify the same participants as requiring further evaluation?
3. What is the accuracy at which each screener classifies the participant as +referral or – referral as measured by the ITALK?

CHAPTER 3
METHODS

Participants

Participants were ten children enrolled in one school from a midsized city in Louisiana. The school was chosen based on the population of Spanish-English bilinguals and ELL preschool and kindergarten students. Bilinguals or ELLs comprised 6% of the total school district population according to data from the 2015 school year. However, the 6% figure does not specify which languages the students spoke. In the school district, females comprised 50% and males 50%, and 81% were considered economically disadvantaged. Within the school-district, the programs provided for children who spoke languages other than English included English-only classroom instruction with English pull-out services for K-5th, and English only classroom instruction for Pre-K. Of the eligible participants, consent forms were returned for 13. Of the initial 13, two moved during the study, and one was excluded due to poor attendance.

Of the ten children who participated in the study, I was able to obtain birthdays for six. For these children, their ages averaged 66.83 mo. (SD = 5.77). Two children were female and eight were males. The children's language profiles were determined using the BIOS. For the BIOS, a caregiver and teacher were interviewed for each child. From these interviews, I determined home and school reported Spanish input/output (SIO) and English input/output (EIO). Following the BIOS manual, I averaged the home and school estimates for each language. The average scores were used to classify each child as either: functional monolingual English (FME; 80% or more English input-output), bilingual English dominant (BED; 60%-80% English input-output), balanced bilingual (BL; 40%-60% input-output for each language), bilingual Spanish dominant (BSD; 60%-80% Spanish input-output), or functional monolingual Spanish (FMS; 80% or more Spanish input-output). Using these procedures, three children were

classified as BED, three were classified as FME, and one was classified as BL. For the three children who I was unable to collect home language data, the teachers reported two as FME and one as BED.

All ten children's English dialect profiles were determined using the DELV-S-I. Recall that the DELV-S-I allows each child's use of English to be classified as reflecting MAE, some variation from MAE, or strong variation from MAE. The DELV-S-I includes five items that target phonology (e.g., voiceless th, voiced th, and consonant cluster ft) and nine that target morphology (e.g., third person subject/verb agreement). Of the participants, five presented an English dialect with some variation from MAE and five presented an English dialect with strong variation from MAE. The DELV-S-I was also used to calculate a nonmainstream dialect density metric (DDM) by dividing the child's nonmainstream responses by the sum of their mainstream and nonmainstream responses. The children's average DDM was .78 (SD =.22; range = .40 to 1.00). Table 3.1 shows the participant's demographic information and their BIOS and DELV-S-I classifications.

Finally, the ITALK questionnaire was used to guide the caregiver and teacher interviews, and these were used to determine each child's referral status. The ITALK asked caregivers and teachers about the child's speech and language in Spanish and English. As mentioned earlier, the ITALK covered five domains: Vocabulary, Speech, Sentence Production, Grammar, and Comprehension, from which a composite score was derived for both the home and school. Following the ITALK manual, the composite score was obtained by adding the individual scores for the five areas and then dividing by five. If a caregiver or teacher stated that they did not know about language use in a category, the score was divided by the number of categories answered.

The questionnaire was completed by the caregivers over the phone in the language of their preference and in person by the teachers.

Table 3.1 Participant description

Participant	Sex	Age (months)	BIOS Spanish input-output score	BIOS English input-output score	BIOS language status	DELV-S-I classification
46	M	66	22	78	BED	Strong variation
60	M	-	0*	100*	FME*	Some variation
61	M	-	6*	94*	FME*	Strong variation
62	M	71	22.5	77.5	BED	Some variation
63	M	60	17.5	83	FME	Some variation
65	M	61	15.5	84.5	FME	Some variation
67	M	-	16.5	83.5	FME	Strong variation
68	M	-	24	76	BED	Strong variation
69	F	75	56*	72*	BED*	Some variation
72	F	68	40.5	59.5	BL	Strong variation

*Teacher report only

Table 3.2 presents the participants' scores across the four domains by setting for the ITALK, and shading is used to highlight each child's best score. As can be seen, teacher data were available for all ten children and caregiver data were available for seven. For those without caregiver data, their best score was determined by the teacher data only.

Table 3.2 ITALK scores

Participant	ITALK English school score	ITALK Spanish school score	ITALK English home score	ITALK Spanish home score
46	3.6*	DK	2	3.4
60	3.6*	DK	-	-
61	2.6*	2.2	-	-
62	4*	DK	3.6	2.6
63	4.8	DK	4.6	5*
65	4.6*	1	4.6*	4.2

(Table continued)

Participant	ITALK English school score	ITALK Spanish school score	ITALK English home score	ITALK Spanish home score
67	3.6	DK	2.6	3.8*
68	4.6	DK	4	5*
69	2.6	3.8*	-	-
72	3.4	4.4*	3.6	4.2

DK= do not know

Recall that the ITALK was used as the reference standard for each child's referral status. On the ITALK and using the child's best language, a score at or below 4.18 in either school or home indicates a speech or language concern. In table 3.2 * indicates the score compared to 4.18 to determine pass/fail. Using the ITALK interviews, six participants failed (46, 60, 61, 62, 67, 69) and were classified as +referral and four passed (63, 65, 68, 72) and were classified as -referral. As it turned out, all children either passed or failed regardless of which set of scores were used (i.e., scores were all above 4.18 or all below). Table 3.3 shows the mean age and standard deviation for the children's ITALK best scores by referral status. The two referral groups did not differ by age as tested by a one-way ANOVA; $F(1,5) = 4.48$, $p = .10$, $\eta^2 = .5$. As expected, ITALK home scores were significantly lower for the +refer than the –refer group as tested by a one-way ANOVA; $F(1,6) = 19.945$, $p = .007$, $\eta^2 = .80$; as were school fail scores $F(1,9) = 6.697$, $p = .032$, $\eta^2 = .46$.

Table 3.3 ITALK scores by referral status

ITALK	Age[a]	ITALK best home score[a]	ITALK best school score
+refer	70.67 (4.509)	3.6 (.20)	3.33 (.59)
-refer	63.00 (4.359)	4.7 (.66)	4.35 (.54)
Total	66.83 (5.776)	4.2 (.66)	3.7 (.78)

[a]Data available for seven participants

Of the children who failed, the BIOS classified two as BED and one as FME and the DELV-S-I identified three of their English dialects as some variation from MAE and three as strong variation from MAE. The -refer group included one BED, two FME, and one BL, of which two produced a dialect with some variation from MAE and two with strong variation from MAE. In other words, the children's language profiles and dialect status did not appear to affect their +/-refer status.

Materials

PLSSST-5. Children completed the PLSSST-5 (Zimmerman et al., 2012a) as described by the manual in approximately 10 minutes. As mentioned earlier, there were six subtests for this screener, but only the language subtest was examined.

The language subtest assessed the following semantic areas: vocabulary, qualitative concepts, quantitative concepts, spatial concepts and time/sequence concepts. Morphology and syntax were also assessed. Items differed based on age. The five-year old items consisted of understanding complex sentences, identifying that which does not belong, object naming, answering hypothetical questions (what do you do if you feel sick?), and sentence repletion. The six-year old items consisted of understanding time/sequence concepts (what did the boy do last?), telling a story in sequence, repairing semantic absurdities (the boy sleeps on a bicycle), and completing similes (if something is very cold, I could say that it is as cold as...). Seven-year old items consisted of understanding quantitative concepts (which girl has the fewest balloons?), making grammatical judgments (Her can eat cookies. Does that sound right or wrong? How would you say it?), using subjunctive mood (only administered in Spanish), deleting sounds (Beat. If I take away the [b] in beat, what word is left?), and formulating sentences with given

words. First, the screener was given in Spanish. Then missed items were administered in English. The criteria to receive a passing score on the language subtest was 4 out of 5.

BESOS. Children completed the BESOS (Lugo-Neris et al., n.d.) according to the manual in approximately 20 minutes. As mentioned earlier, the BESOS included four subtests. For the semantic subtests, there were 10 items for four-year olds and 12 items for 5-to 6-year-olds. These items evaluated the child's understanding of categories or concepts (e.g., tell me all the foods you can think of, and red, blue, yellow and green are all…). Responses were allowed in either language. The morpho-syntax subtest included 16 items for four-year-olds and 17 items for 5- to 6-year olds. These items used cloze and sentence repetition in each language to elicit English past tense, third person present tense, and Spanish articles, direct object clitics, and subjunctives. All children were given the English and Spanish morpho-syntax subtests. Based on the findings by Lugo-Neris et al. (2015), only the best score on each subtest was used to determine whether a child passed or failed the screener. The morpho-syntax subtest contained 16 (age 4) or 17 (age 5-6) items. The semantic subtest consisted of 10 (age 4) or 12 (age 5-6) items. Following the scoring guide of the BESOS (E. Peña, personal communication, June 13, 2017), raw scores for each language were compared and the higher scores were used to select the best language score. These raw scores were then used to calculate a standard score using the following formula: ((raw score – mean for that age group) / standard deviation for that group) X 15 + 100. The cut scores for pass varied depending on the subtest, language examined, and child's age.

DELV-S-II. Children completed the DELV-S-II (Seymour et al., 2003) as described by the manual in approximately 20 minutes. The DELV-S-II included 17 items. Items 1 through 3 and 8 through 11 targeted morphology (e.g., copula verb, auxiliary verb and possessive

pronouns). For these items, each child was shown a picture and asked to finish a sentence or answer a question about the picture. Items 4 through 7 measured the child's ability to answer wh-questions. Items 12 through 17 consisted of non-word repetitions. The total number of incorrect responses were summed to determine the child's diagnostic risk status. Scores falling below the highest risk category were used to determine who passed.

Procedures

The study was approved by Louisiana State University's Institutional Review Board. Consent forms were sent to all students who may have been bilingual or ELL as determined by their teachers. All Latino children who spoke Spanish, English or both and who did not have a significant disability by teacher report were asked to participate. After agreeing to participate, all participants were administered the same three screeners, PLSSST-5, BESOS, and DELV-S-II. The order of the screeners was counter-balanced, with one participant receiving order 1 (BESOS Spanish, PLSSST-5, DELV-S-II, BESOS English), three participants receiving order 2 (PLSSST-5, BESOS Spanish, DELV-S-II, BESOS English), three participants receiving order 3 (BESOS Spanish, PLSSST-5, BESOS English, DELV-S-II), and three participants receiving order 4 (PLSSST-5, BESOS Spanish, BESOS English, DELV-S-II). Screener order was selected at the beginning of the study, and due to participant attrition only one child received order 1.

A bilingual graduate student administered the screening instruments. Administration took place at each participant's school over two 30-minute periods. Data was collected and recorded using a digital voice recorder. Numbers and alpha-numeric codes were assigned to the participants' data to maintain confidentiality.

Reliability

To assess reliability of scoring and data coding, four of the participants' screening data were randomly selected. A second Spanish-English bilingual graduate student listened to recorded responses for two of the selected participants and scored them independently on a blank screening protocol. Additionally, written responses on each screener form for the other two participants were scored by the second graduate student. There were twenty opportunities to agree on the PLSSST-5, 224 on the BESOS, and 68 on the DELV-S-II. All four participants' responses were examined manually. Agreement between the four sets of scores was 96%.

CHAPTER 4
RESULTS

PLSSST-5, BESOS, and DELV-S-II Pass Rates

The first analysis examined the percentage of participants who passed the PLSSST-5, BESOS, and DELV-S-II. Table 4.1 lists the participants' screening outcomes for each tool (see appendix for each child's raw data). Distribution of the pass/fail rates for the individual screeners is shown in Table 4.2. Results indicated that the PLSSST-5 and DELV-S-II passed (30%) and failed (70%) the same proportion of participants. The BESOS identified the highest portion of the participants as passing (60%) and the lowest proportion as failing (40%).

Table 4.1 Pass/fail outcomes per screener

Participant	PLSSST-5	BESOS	DELV-S-II
46	Fail	Pass	Fail
60	Fail	Pass	Fail
61	Fail	Pass	Pass
62	Fail	Fail	Fail
63	Pass	Pass	Fail
65	Fail	Fail	Pass
67	Fail	Fail	Fail
68	Pass	Pass	Pass
69	Fail	Fail	Fail
72	Pass	Pass	Fail

Table 4.2 Screener pass rates

	Pass		**Fail**	
Screener	**N**	**%**	**N**	**%**
PLSSST-5	3	30	7	70
BESOS	6	60	4	40
DELV-S-II	3	30	7	70

Pass rates by the participants' assigned order, language status, and dialect status are presented in Tables 4.3, 4.4, and 4.5. For all screeners, neither assigned order, language status, or dialect status led to different pass rates as tested by χ^2 ($p > .05$).

Table 4.3 Pass rates by test order

	Test order			
	1	2	3	4
PLSSST-5	100%	33%	0%	33%
BESOS	100%	67%	33%	67%
DELV-S-II	0%	33%	33%	33%
n	1	3	3	3

Table 4.4 Pass rates by language status

Screener	BED	FME	BL
PLSSST-5	25%	20%	100%
BESOS	50%	60%	100%
DLEV-S-II	25%	40%	0%
n	4	5	1

Table 4.5 Pass rates by dialect status

Screener	Some variation	Strong variation
PLSSST-5	20%	40%
BESOS	40%	80%
DLEV-S-II	20%	40%
n	5	5

Screening Outcomes Across PLSSST-5, BESOS, and DELV-S-II

Table 4.6 presents the pass rates for the screeners combined. One participant passed all three screeners, three passed only one screener, and three failed all three.

Table 4.6 Screener combination pass rates

Combination of screeners	**N**	**%**
Passed all screeners	1	10
Passed only 1 screener	3	30
Failed all screeners	3	30
PLSSST-5 & BESOS	2	20
PLSSST-5 & DELV-S-II	0	0
BESOS & DELV-S-II	1	10

Screening Outcomes Compared to ITALK

The results of the screeners were then compared to the caregiver and teacher ratings on the ITALK. This comparison serves to examine the classification accuracy of the screeners using the ITALK as the reference standard for the participants' referral status (see Table 4.7). Recall

that sensitivity is the proportion of participants who score at risk for language impairment and who actually are at risk, and specificity is the proportion who are not at risk and score as such. The PLSSST-5 yielded the highest sensitivity rate (100%), followed by the DELV-S-II (86%), and then the BESOS (50%). The PLSSST-5 and the BESOS yielded the highest specificity rates (75%), and the DELV-S-II yielded the lowest (50%).

Table 4.7 Screener combination pass rates

ITALK	PLSSST-5	BESOS	DELV-S-II
+referral			
46	Fail	Pass	Fail
60	Fail	Pass	Fail
61	Fail	Pass	Pass
62	Fail	Fail	Fail
67	Fail	Fail	Fail
69	Fail	Fail	Fail
Sensitivity	100%	50%	86%
-referral			
63	Pass	Pass	Fail
65	Fail	Fail	Pass
68	Pass	Pass	Pass
72	Pass	Pass	Fail
Specificity	75%	75%	50%

Table 4.8 lists the false positive and false negative rates. False negatives refer to those who passed the screening but were identified by the ITALK as needing further assessment and false positives refer to those who the screeners failed but were typically developing. Results indicate that the BESOS had the highest rate of erroneous classification (40%) when compared to the ITALK. The PLSSST-5 had the highest accuracy (90%) followed by the DELV-S-II (70%).

Table 4.8 False positive/negative rates

	PLSSST-5	BESOS	DELV-S-II
False positive	1 (10%)	1 (10%)	2 (20%)
False negative	0	3 (30%)	1 (10%)
Overall Accuracy	90%	60%	70%

CHAPTER 5
DISCUSSION

The purpose this study was to learn more about language screening instruments by examining the pass rates of the PLSSST-5, BESOS, and DELV-S-II when given to a group of bilingual and/or ELL children. Screeners were chosen based on their norming sample. The PLSSST-5 is normed on monolingual Spanish speakers and Spanish dominant bilinguals, the BESOS is normed on Spanish-English bilingual children with a broad range of language abilities, and the DELV-S-II includes a small group of Spanish-influenced English speakers.

First Research Question

The first question addressed the proportion of participants who passed the three screeners. Of the three screeners evaluated, the highest proportion of participants (60%) passed the BESOS. The PLSSST-5 and the DELV-S-II passed the same proportion of participants (30%).

Second Research Question

The three screeners disagreed not only in pass rates but also in the specific children who passed. Although the PLSSST-5 and DELV-S-II identified the same proportion of participants as requiring further referral, only one of the participants passed both of the screeners. This same participant is the only one that passed all three screeners (10%). Overall, 30% of the participants failed all the screeners, and 60% inconsistently passed or failed the screeners. The results of the current study suggest that the three screeners have poor consistency between them.

Third Research Question

The ITALK was used as the reference standard to evaluate the classification accuracy of the screeners. The PLSSST-5 had the highest sensitivity rate (100%), correctly classifying all those identified by the ITALK as +referral. This was followed by the DELV-S-II (86%) and then the BESOS (50%). Based on these results, it appears that BESOS was least able to identify

participants who required further evaluation. The DELV-S-II had the lowest specificity rate (50%). Therefore, it is more likely than the others to result in an over referral of participants. Recall that the DELV-S-II is not normed on bilingual speakers but on Spanish-influenced English speakers. It could be possible that bilingual language differences could account for low specificity rates on this screener and that the participants' English ability could be related to the high rate of over referral. However, when looking at the participants' language classification or dialect, there were no clear distinctions between the participants' abilities in each language and their results on the DELV-S-II or on the other screeners.

Another way to examine the classification accuracy of the screeners is by false positive and false negative rates; both are considered errors. Preferably, a screener that results in a high rate of accurate identification with low rates of inaccurate identification is ideal. The 10% overall error rate of the PLSSST-5 included one false positive. In the current study, the PLSSST-5 more accurately classified children than the other two screeners. The overall error rate for the BESOS (40%) and the DELV-S-II (30%) included both false positives and false negatives. The false positive rate for the DELV-S-II (20%) was higher than for the BESOS (10%). The false negative rate for the BESOS (30%) was higher than for the DELV-S-II (10%). None of the screeners had complete agreement with the ITALK. The high false negative rate of the BESOS is of concern. If errors occur, it is preferable that they occur as false positives. False positives when screening should receive additional testing, and a language impairment can be ruled out at that time. False negatives, on the other hand, are those who require further evaluation but they are never referred.

Findings as Related to Previous Studies

The results of this study suggest that the PLSSST-5, BESOS, and DELV-S-II have poor consistency in their identification of participants as +referral and –referral. This is in line with

findings from Blaxely et al. (1983) who found that screeners differed in the children identified as passing or failing. Additionally, Moland (2011) found that a large percentage of her participants inconsistently passed or failed the screeners utilized in her study. Illerbrun et al. (1985) contend that there are several factors that can affect consistency among screeners. One of those factors is the language skills that each screener measures (i.e. semantics, morphology, receptive language, etc.). As shown in the current study, screening outcomes can be further complicated by the actual language and dialect targeted within the tool (e.g. Spanish, some variation of MAE). Differences in the norming samples of tools may also affect screening outcomes. The PLSSST-5 is normed predominantly on Spanish speakers while the DELV-S-II is normed on Spanish-influenced English speakers.

Lugo-Neris et al. (2015) and Kilgus et al. (2014) argue that because screeners do not automatically result in a diagnosis, there can be some leniency in sensitivity and specificity rates. They advocate for sensitivity rates of 80% or higher and specificity rates of 70% or higher. Using sensitivity rates of 80% or higher and specificity rates of 70% or higher as a guideline, the PLSSST-5 produced adequate sensitivity and specificity rates. The BESOS yielded poor sensitivity but adequate specificity. The DELV-S-II's sensitivity was adequate but specificity was not. In the studies discussed in the literature review, sensitivity rates of screeners were also lower than specificity rates. Across studies, these findings indicate that current screeners are more likely to overlook those in need of further evaluation. Given this, more research is needed to evaluate the screeners examined in the current study.

Limitations

This study had several limitations. First, the study was conducted with a small number of participants from the same school. This led the study to be predisposed to spectrum bias.

According to Dollaghan and Horner (2011) "spectrum bias is a concern when diagnostic accuracy is calculated from a sample of participants who do not represent the full characteristics that would be encountered when the index measure is used in a real-world clinical context" (p. 1078). Future studies should include a larger group and a more diverse group of bilingual and ELL participants. Second, data were missing for four of the participants and this included their date of birth, maternal education, and caregiver responses on the BESOS and ITALK. Third, the ITALK was used as the reference standard. In support of the ITALK, the teacher and caregiver interviews collected in the current study had a very high agreement rate on –referral and +referral. However, Rodriguez and Guiberson (2011) and Gregory and Oetting (under review) included the TROLL in their studies which is also a teacher rating scale. Both of these research teams used the teacher ratings as a screening measure rather than as a reference standard. Both of these studies found that the TROLL led to low classification accuracies. While the ITALK provides important information about teacher and caregiver perceptions, it consists of self-reported data and cannot be independently verified. To test the validity of the ITALK, the results should be compared to an actual diagnostic measure. Fourth, in the current study the same examiner administered the reference standard and all the screeners. The examiner also had information regarding teacher and caregiver concerns about the participants' speech and language performance. Within the diagnostic accuracy literature, this method of data collection increases the likelihood of the results being affected by a subjectivity bias.

Clinical Implications

Findings from the current study indicate that of the PLSSST-5, the BESOS, and DELV-S-II, no screener emerged as an ideal instrument. Although the PLSSST-5 had the highest accuracy rate, these findings need to be replicated with a larger sample and other reference

standards. Until this work is done, the three screeners examined should be supplemented with additional tools. Use of these instruments as a single measure may result in under or over identification of those requiring further evaluation.

Future Directions

Future studies should continue to compare the results of screeners, teacher and caregiver rating scales, and diagnostic instruments to determine if there is a combination that helps identify those needing further assessment more efficiently. Furthermore, instruments that attempt to quantify a child's language ability and dialect should continue to be explored on how they relate to performance on a screener and on teacher and caregiver rating scales. Another area of focus should analyze performance on individual test items to try and explain why results vary among screeners.

Conclusion

The current study found that the pass rates of the three screeners varied and that they passed different children. While the PLSSST-5 provided the highest accuracy rating when compared to the reference standard, the small sample size and lack of comparison to a diagnostic assessment reduce the certainty of the results. The ITALK, which was used as the reference standard was based on teacher and caregiver ratings, and these ratings were highly consistent with each other. However, since the ITALK results were not compared to a diagnostic assessment, results should be taken with caution. Overall, the results suggest that screeners should not be used as stand-alone tools to decide if children should be referred to a speech-language pathologist for an evaluation.

REFERENCES

Alcoff, L. M. (2005). Latino vs. Hispanic: The politics of ethnic names. *Philosophy & Social Criticism*, *31*(4), 395-407.

Amaro, H., & Zambrana, R. E. (2000). Criollo, mestizo, mulato, LatiNegro, indígena, white, or black? The US Hispanic/Latino population and multiple responses in the 2000 census. *American Journal of Public Health*, *90*(11), 1724–1727.

Bayley, R. (1994). Consonant cluster reduction in Tejano English. *Language Variation and Change*, *6*(03), 303-326.

Baker, G. (1975). Social functions of language in a Mexican-American community. *American Anthropologist,* 76(2), 457-458.

Bankson, N. W. (1990). *Bankson Language Test-2.* Austin, TX: Pro-Ed.

Baxley, L., Clinker, M., & Warr-Leeper, G. (1983). Two language screening tests compared with developmental sentence scoring. *Language, Speech and Hearing Services in Schools, 14*(1), 38-46.

Bedore, L. M., Pena, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K. A. I., Bohman, T.M., & 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.

Berk-Seligson, S. (1980). A sociolinguistic view of the Mexican-American speech community: A review of the literature. *Latin American Research Review, 15*(2), 65-110.

Bishop, D.V.M. (2006), CCC-2; *Children's Communication Checklist-2, United States Edition.* San Antonio, TX: Pearson.

Carrow-Woolfolk, E. (1985). *Test for Auditory Comprehension of Language-Revised*. Allen, TX: DLM Teaching Resources.

Census.gov. (2014.). Retrieved May 23, 2016, from http://www.census.gov/

Craig, H. K., Washington, J. A., & Thompson-Porter, C. (1998). Average C-unit lengths in the discourse of African American children from low-income, urban homes. *Journal of Speech, Language, and Hearing Research*, *41*, 433–444.

Craig, H. K. & Washington, J. A. (2000). An assessment battery for identifying language impairments in African American children. *Journal of Speech, Language, and Hearing Research*, *43*, 366-379.

Craig, H. K. & Washington, J. A. (2002). Oral language expectations for African American preschoolers and kindergartners. *American Journal of Speech-Language Pathology*, *11*, 59-70.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1077-1088.

Dunn, L. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test.-IV*. San Antonio, TX: Pearson.

Dickinson, D., McCabe, A., & Sprague, K. (2001). *Teacher rating of oral language and literacy (TROLL): A research-based tool.* Ann Arbor, MI: Center for the Improvement of Early Reading Achievement, University of Michigan.

Fluharty, N. B. (1978). *Fluharty Preschool Speech and Language Screening Test*. Allen, TX: DLM Teaching Resources.

Fluharty, N.B. (2001). *Fluharty Preschool Speech and Language Screening Test*. (*2nd ed).* Austin: ProED.

Fought, C. (2006). Talkin'with mi gente (Chicano English). *American voices: How dialects differ from coast to coast.* Malden, MA: *Blackwell Publishing*, 233-237.

García, O., & Kleifgen, J. A. (2010). *Educating emergent bilinguals: Policies, programs, and practices for English language learners*. Teachers College Press.

Gauthier, S. V., & Madison, C. L. (1998). *Kindergarten Language Screening Test-Second Edition*. Austin, TX: Pro-Ed.

Good, R.H., Gruba, J., & Kaminski, R.A. (2009). *Dynamic Indicators of Basic Early Literacy Skills Next.* Longmont, CO: Cambium Learning Group.

Gonzales, M. (1999). Crossing social and cultural borders. In *Speaking Chicana* (pp. 13-38). Tucson: The University of Arizona Press.

Gregory, K. D., & Oetting, J.B. (2017). *Use of Teacher Ratings when Screening Nonmainstream English-Speaking Kindergarteners for Language Impairment.* Manuscript submitted for publication.

Gutiérrez-Clellen, V. F.,& Kreiter, J., (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24*, 267-288.

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., & Wagner, C. (2007). Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Applied Psycholinguistics, 29*(01).

Hayes-Bautista, D. E., & Chapa, J. (1987). Latino terminology: Conceptual bases for standardized terminology. *American Journal of Public Health*, *77*(1), 61–68.

Hosp, J.L., & Reschly, D.J. (2003). Referral rates for intervention or assessment: A meta-analysis of racial differences. *The Journal of Special Education, 37*(2), 67-80.

Illerbrun, D., Haines, L., & Grenough, P. (1985). Language identification screening test for kindergarten: A comparison with four screening and three diagnostic language tests. *Language, Speech, and Hearing Services in Schools*, *16*(4), 280-292.

Illerbrun, D., McLeod, J., Greenough, P., & Haines, L. (1984). *Language Identification Screening Test for Kindergarten*. Saskatoon, Saskatchewan: EDEXC Occasional Publications, University of Saskatchewan.

Kilgus, S. P., Methe, S. A., Maggin, D. M., & Tomasula, J. L. (2014). Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of evidence supporting use in universal screening. *Journal of school psychology*, *52*(4), 377-405.

Kracht, M., & Klein, U. (2014). The grammar of code-switching. *Journal of Logic, Language and Information, 23*(3), 313-329.

Lee, L. L., and Canter, S. M. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*, 315-340.

Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., & Gillam, R. B. (n.d.). *Bilingual English Spanish Oral Screener.* Not yet published.

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.

Moland, C. M. (2011). *Comparison of the Diagnostic Evaluation of Language Variation-Screening (DELV-ST) to two other screeners for low-income, African American children* (Unpublished doctoral dissertation). Louisiana State University, Baton Rouge, Louisiana.

Newcomer, P., & Hammill. D. (1982). *Test of Language Development Primary*. Austin, TX: Pro-Ed.

Peña, E.D., Gutiérrez-Clellen, V.F., Iglesias, A., Goldstein, B.A., & Bedore, L.M. (2014). *Bilingual Input-Output Survey*. San Diego, CA: AR-Clinical Publications.

Peña, E.D., Gutiérrez-Clellen, V.F., Iglesias, A., Goldstein, B.A., & Bedore, L.M. (2014). *Instrument to Assess Language Knowledge*. San Diego, CA: AR-Clinical Publications.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, *25*, 15-24.

Paradis, J., Genesee, F., & Crago, M.B. (2011). *Dual Language Development and Disorders*. Baltimore: Paul H. Brooks Publishing Co.

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

Rodriguez, B.L., & Guiberson, M. (2011). Using a teacher rating scale for language and literacy skills with preschool children of English-speaking, Spanish-speaking, and Bilingual backgrounds. *Early Childhood Education Journal,* 39 (5), 303-311.

Santa Ana A., O. (1993). Chicano English and the nature of the Chicano language setting. *Hispanic Journal of Behavioral Sciences, 15*(1), 3-35.

Semel, E. M., & Wiig, E. H. (1980). *Clinical Evaluation of Language Functions-Elementary Screening Test*. Columbus, OH: Charles E. Merrill.

Seymour, H.N., Roeper, T.W., de Villiers, J., & de Villiers, P.A. (2003). *Diagnostic Evaluation of Language Variation-Screening Test*. *San Antonio, TX*: PsychCorp.

Seymour, H. N., Roeper, T. W., & de Villiers, J. (2005). *Diagnostic Evaluation of Language Variation-Norm Referenced*. San Antonio, TX: The Psychological Corporation.

Simon-Cereijido, G., & Gutierrez-Clellen, V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics*, *28*(2), 317-339.

Valdés, G., & Figeroa, R.A., (1994). *Bilingualism and testing: A special case of bias*. *Norwood, NJ*: Ablex

Washington, J. A., & Craig, H. K. (2004). A language screening protocol for use with young African American children in urban settings. *American Journal of Speech-Language Pathology, 13*, 329-340.

Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: Why variations in bilingual experiences matter. *Frontiers in psychology*, *6*.

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

Zimmerman, I.L., Steiner, V.G., & Evatt Pond, R. (2012a). *Preschool Language Scales Spanish Screening Test-fifth Edition*. Bloomington, MN: NCS Pearson.

Zimmerman, I.L., Steiner, V.G., & Evatt Pond, R. (2012b). *Preschool Language Scales-fifth Edition*. Bloomington, MN: NCS Pearson.

Zimmerman, I.L., Steiner, V.G., & Evatt Pond, R. (2012c). *Preschool Language Scales Spanish-fifth Edition*. Bloomington, MN: NCS Pearson.

APPENDIX
RESULTS FOR EACH SCREENER BY PARTICIPANT

PLSSST-5: Individual Scores

Participant	Score	Pass/Fail
46	3	Fail
60	3	Fail
61	2	Fail
62	2	Fail
63	5	Pass
65	2	Fail
67	3	Fail
68	5	Pass
69	3	Fail
72	5	Pass

BESOS Semantics: Individual Scores

Participant	Spanish semantics raw score	Spanish semantics standard score	Spanish semantics pass/fail	English semantics raw score	English semantics standard score	English semantics pass/fail
46	6	85	Fail	9	113	Pass
60	7	92	Pass	10	120	Pass
61	11	104	Pass	9	101	Pass
62	1	51	Fail	2	63	Fail
63	8	103	Pass	8	108	Pass
65	1	55	Fail	1	61	Fail
67	5	78	Fail	8	100	Pass
68	12	114	Pass	7	93	Pass
69	0	52	Fail	2	59	Fail
72	9	106	Pass	7	99	Pass

BESOS Morpho-syntax: Individual Scores

Participant	Spanish morpho-syntax raw score	Spanish morpho-syntax standard score	Spanish morpho-syntax pass/fail	English morpho-syntax raw score	English morpho-syntax standard score	English morpho-syntax pass/fail
46	2	68	Fail	9	98	Pass
60	5	79	Fail	9	98	Pass
61	6	69	Fail	13	100	Pass
62	1	64	Fail	5	84	Fail
63	13	112	Pass	6	92	Pass
65	3	76	Fail	9	102	Pass
67	6	69	Fail	8	75	Fail
68	14	105	Pass	16	115	Pass
69	1	47	Fail	1	40	Fail
72	8	90	Pass	8	94	Pass

DELV-S-II: Individual Scores

Participant	Risk error score	Overall/pass/fail
46	9	Fail
60	9	Fail
61	5	Pass
62	18	Fail
63	9	Fail
65	7	Pass
67	9	Fail
68	6	Pass
69	17	Fail
72	9	Fail

VITA

Mariajosé Bosanko graduated from Kendall College of Art and Design in 2002 with a Bachelor of Fine Arts degree in furniture design. After working in the public school system, she completed a Master of Arts in Language, Literacy, & Sociocultural Studies with an emphasis in bilingual education from the University of New Mexico in 2008. She continued to work in the field of education until 2015 when she began pursuing a Master of Arts degree in Communication Science and Disorders at Louisiana State University. She is continuing her studies and anticipates obtaining a Ph.D. in Communication Science and Disorders from Louisiana State University in 2021.