Establishing the Preliminary Validity and Reliability Evidence of the Spanish Screener for Language Impairment in Children (SSLIC)

by

Brandon Earl Smith

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2023 by the
Graduate Supervisory Committee:

M. Adelaida Restrepo, Co-Chair
Shelley Gray, Co-Chair
Jean Brown

ARIZONA STATE UNIVERSITY

May 2023

ABSTRACT

Currently, there are few quality tools available to screen for developmental language disorder (DLD) in Spanish-speaking children despite the continued increase of this population in the United States. The lack of valid and reliable screening tools may be a factor leading to difficulties with the identification of and delivery of services to these children. This study plans to improve upon the screening of Spanish-English bilingual children.

The Spanish Screener for Language Impairment in Children (SSLIC) tests Spanish oral language skills in Spanish-speaking children. It measures language skills through morphology elicitation of Spanish clitics, prepositions, derivational morphemes, subjunctive verb tenses, and articles and repetition of nonwords and sentences, which have all been shown to be affected in Spanish-speaking children with DLD.

The purpose of the study is to provide preliminary validity evidence of the SSLIC. Children's results on the SSLIC were compared to other validated measures. Fourteen Spanish-English bilingual students were recruited: 11 children with typical language development (TD) and 3 with DLD. The *Bilingual English-Spanish Assessment* and the *Dynamic Measure of Oral Narrative Discourse* were used to establish preliminary validity evidence. Pearson correlations were run to determine if SSLIC scores correlated with other validated measures. Significant correlations were found between the SSLIC's scores and scores on the BESA. One-way analysis of variance (ANOVA) was used to determine mean differences between groups. No significant mean differences for SSLIC scores were found between children with typical and atypical language. Yet, effect sizes suggested group differences. Point to point analysis revealed that the SSLIC has excellent inter-rater reliability.

Despite a small sample size, this study serves as preliminary evidence that the SSLIC is both valid and reliable and supports that the SSLIC has the potential to be used

as a screening tool for DLD for Spanish-speaking kindergarten and 1st grade students with further validation, which should continue.

TABLE OF CONTENTS

Introduction

Of all the children enrolled in the US public schools, roughly 10.2 percent of them are bilingual. Of those, approximately 75.2% speak Spanish as their native language (US Department of Education, 2018). In the United States as a whole, there are about 62 million individuals who identify as Hispanic and Latino (18.7% of the total population) and about 13.2% of the United States' population speaks Spanish at home (United States Census Bureau, 2021). Between fall 2009 and fall 2019, the number of Hispanic students in US public schools increased from 11 million to 14.1 million before decreasing to 13.8 million in fall 2020, which means this population grew from 22 percent of the population to 28 percent of the population (National Center for Education Statistics, 2022). Additionally, it is predicted that these numbers will continue to increase (National Center for Education Statistics, 2022).

Speech-language pathologists working in public schools must be prepared and have the tools to work with children from culturally/linguistically diverse backgrounds who have developmental language disorder (DLD)[1]. Tomblin (1997) found that 7.4% of kindergarten children who were screened and tested from a sample of 7,218 children had a developmental language disorder. Therefore, we expect similar rates of bilingual students with language disorders.

The Spanish Screener for Language Impairment in Children (SSLIC) was developed to identify children who may be at risk of a language impairment and require further language assessment given the limited number of validated measures to accurately screen this group of children. Currently, there are very few screening and assessment

[1] The terms SLI, PLI and LI were used previously in research. There is a consensus now to use DLD, which we use to describe this population (Bishop, 2017).

tools to identify children who would benefit from language-based speech-language pathology services. Molanda and Oetting (2021), described how there is a great need for quality screening tools in the field of speech-language pathology that are accurate and valid for that purpose. They compared three screening tools and showed that many screeners have "high fail rates and inconsistent outcomes" (Molanda & Oetting, 2021). Although their article focused on children who speak African American English, their research findings and suggestions can be applied to bilingual students in the US. Hispanic children, in particular, are 46% less likely to receive speech or language services when compared to their non-hispanic white peers (Morgan et al., 2017). It's hypothesized that this could be due to inappropriate diagnoses of language difference, rather than language disorder. Moreover, children from homes that don't speak English were found to receive speech and language services at a rate that was 50% lower when compared to their primarily English-speaking peers (Morgan et al., 2017). Yamasaki and Luk (2018) found that English proficient bilinguals were underidentified for communication disorders in grades 3, 4, and 5 and emerging bilinguals were overidentified for communication disorders in grades 4 and 5. These studies highlight the inconsistencies that currently exist with the identification of DLD in bilingual children, possibly due to the limited number of validated measures available. The current research study plans to improve on the assessment of Latino Spanish-English bilingual students.

The present recommendation for assessing bilingual students is to gather and combine information from different sources and measures to differentiate those with DLD from those without DLD (Castilla, et al, 2020; Dollaghan & Horner, 2011; Paradis et al., 2013). In their study, Paradis et al. (2013) administered standardized assessments in

English that tested “nonword repetition, tense morphology, narrative story grammar, and receptive vocabulary.” The results from these tasks, besides vocabulary, aided in determining the presence of DLD. Yet, a parent questionnaire that was used provided even more information that led to the differential diagnosis of DLD. Further, Castilla-Earls et al. (2020) argued that combining information from “language experience questionnaires, bilingual language sample analysis using large-scale reference databases, evaluation of learning potential, and standardized testing” (p. 1116) can provide converging evidence about a child’s language abilities in various contexts and ultimately a differential diagnosis of DLD.

Questionnaires completed by a child’s parents and teachers have been shown to accurately represent a child’s linguistic abilities and are recommended for use when evaluating the language skills of DLL students (Paradis et al., 2013; Restrepo, 1998). Collecting a language sample and comparing the grammatical accuracy and sentence complexity using a database such as the Systematic Analysis of Language Transcripts (SALT) gives clinicians quantitative data regarding the expressive language skills of a child and compares that data to their peers, which can then inform clinical decision-making (Pezold et al., 2020). Dynamic assessments focus on a child’s learning potential and use a test-teach-retest model to judge a child's ability to learn and apply new linguistic information (Orellana, Wada, & Gillam, 2019). Generally, children who show little improvement despite instruction and score poorly on both the pre- and posttests are more likely to be categorized as DLD (Orellan, Wada, & Gillam, 2019). Those assessments are appropriate for full evaluation. However, we still need a validated measure to screen bilingual children.

Standardized assessments can provide insight into a child's language abilities across multiple domains, are useful for comparing the language abilities of children to their same-aged peers and backgrounds, and are often used to qualify children for services in a school-based setting (Karem and Washington, 2021). However, they don't always provide the most accurate information about a child's linguistic abilities, especially when an assessment doesn't account for variables such as dual language learning and/or cultural differences or was not normed on bilingual populations (Lazewnik et al., 2019). Few assessments consider the bilingual factors. For example, the BESA was normed on bilingual populations and uses a "language index score" using the best score in either Spanish or English for both the semantics and morphosyntax subtests and then combining them to generate a new score. This measure has been validated providing the the following sensitivity and specificity:

- 4-year-olds: 92% sensitivity and 86% specificity. (Peña et al., 2018).
    - Cut score = 86
- 5-year-olds: 89% sensitivity and 85% specificity. (Peña et al., 2018).
    - Cut score = 86
- 6-year-olds: 96% sensitivity and 92% specificity (Peña et al., 2018).
    - Cut score = 81

In addition, when the information gained from this assessment is combined with information gained from a language sample, the overall accuracy of diagnosis and identification of DLD is improved (Lazewnik et al., 2019). Therefore, there is a great need for appropriate screening measures to initiate and evaluate the need for further

assessment. For the purpose of this study, various assessment tasks will be used to determine the validity and reliability of the SSLIC.

There are few quality tools available to screen for language impairment in children who primarily speak Spanish. Lugo-Neris et. al. (2015) discussed how a bilingual screening tool for language impairment in children can be useful to predict DLD in bilingual preschoolers. In particular, the Bilingual English Spanish Oral Screener (BESOS) was able to predict DLD in bilingual preschoolers with an overall accuracy of 81% when evaluating children in both Spanish and English (Lugo-Neris et al., 2015). It was found to have a sensitivity of 95.2% and a specificity of 71.4%, with an overall accuracy of 81% for predicting risk of DLD (Lugo-Neris et al., 2015), which is considered "fair" at accurately discriminating between DLD and TD (Plante & Vance, 1994). However, this measure is not yet validated nor available commercially to clinicians. The Preschool Language Scales - 5th Edition Spanish (PLS-5S) also offers a screening tool for Spanish-speaking children. The comprehensive PLS-5S assessment reports the sensitivity for identifying DLD to be .85 and specificity for identifying TD to be .88, which means that the assessment is "fair" at identifying both Spanish-speaking children with language disorders and children without language disorders (Leaders Project, 2013), although this has not been validated outside the test developers' group. This, therefore, makes the PLS-5S's construct validity, the ability to evaluate what it intends to evaluate, insufficient (Leaders Project, 2013). Test-retest reliability and inter-item consistency were also found to be insufficient (Leaders Project, 2013). Inter-examiner reliability was found to be adequate, but the test reviewers speculate that this measure would likely be lower in practice considering that the original examiners

received special training when scoring the assessment (Leaders Project, 2013). Ultimately, the validity and reliability of the comprehensive PLS-5S screener is questionable given that there is little research available.

The Spanish Screener for Language Impairment in Children (SSLIC) currently screens a child's language abilities by measuring the appropriate use of the following grammatical structures: Spanish clitics, prepositions, derivational morphemes, subjunctive verb tenses, and articles. Areas sensitive to DLD in Spanish-speaking and Spanish-English bilingual children (Castilla et al., 2016; Castilla et al., 2021; Kapantzoglou et al., 2016; Morgan et al., 2013; Muñoz & Brimo, 2017). It also includes a sentence repetition task and a nonword repetition task, which rely on a child's ability to, auditorily perceive language, encode and assemble phonological stimuli, memorize, execute a motor plan, and have also been found to be good measures to identify DLD (Archibald & Joanisse, 2009; Armon-Lotem & Meir, 2016; Estes et al., 2017; Girbau & Schwartz, 2007; Kapantzoglou et al., 2016; Muñoz & Brimo, 2017; Ortiz, 2021; Schwob et al., 2021). It is now crucial to determine the validity and reliability of these measures to determine the overall validity and reliability of the SSLIC. The measure was originally validated with the Clinical Evaluation of Language Fundamentals - Fourth Edition, Spanish (CELF-4 Spanish); however, upon further examination the cut scores on the CELF-4 Spanish were not appropriate for the Arizona sample as they overidentified children as having DLD (Barragan et al, 2018; Restrepo et al., 2010).

The SSLIC measure is based on the language characteristics of Spanish-speaking children with DLD. Specifically, it tests Spanish clitics, prepositions, derivational morphemes, subjunctive verb tenses, and articles to make inferences about a bilingual

Spanish-English child's language abilities (Castilla et al., 2016; Castilla et al., 2021; Kapantzoglou et al., 2016; Morgan et al., 2013; Muñoz & Brimo, 2017). Castilla-Earls et al. (2021) found that tasks requiring the appropriate use of articles, clitics, adjectives, verbs, and the subjunctive mood showed a difference between children with typical language skills and children with developmental language disorders. Of these grammatical structures, tasks that tested a combination of verbs and the subjunctive mood supported diagnostic accuracy the best. These tasks yielded "acceptable sensitivity and good specificity" (Castilla-Earls et al., 2021). Moreover, assessing a child's use of direct object pronouns (aka Spanish clitics) and articles have also been found to differentiate typical and DLD Spanish-speaking children (Castilla Earls et al., 2015; Morgan et al., 2013). Testing either of these will show differences between typical language development and DLD. Of the two, it's reported that articles tend to be a stronger predictor of language impairment, especially when the child is considered a balanced bilingual. These research findings support the need to analyze the predictive power of both of these measures.

Nonword repetition has been shown to be effective in identifying individuals with DLD (Estes et al., 2007; Girbau & Schwartz, 2007; Ortiz, 2021; Schwob et al., 2021). Both monolingual and bilingual children with DLD are less accurate at repeating nonsense words (Schwob et al., 2021). Although both typically developing children and children with DLD performed worse as syllable length increased, children with DLD performed worse across all syllable lengths (Schwob et al., 2021). Girabau and Schwartz (2007) and Ortiz (2021) speculated that a nonword repetition could be valuable when used as a screening tool to determine if a bilingual child presents with a language

disorder or difference. However, it would have to be tested on more children to generalize their findings. Additionally, Schwob et al. (2021) and Ortiz (2021) recommend using "quasi-universal tasks" when creating a list of nonwords to use with bilingual children as this will improve diagnostic accuracy by decreasing the phonotactic constraints present in some languages and not others.

Archibald and Joanisse's (2009) findings contradict previous research on the predictive power of nonword repetition tasks. They examined the effectiveness of both nonword repetition and sentence repetition tasks when they are used to identify children with DLD, working memory impairments, or both of these impairments. They found that sentence repetition tasks scored at the item level, i.e. sentence by sentence, were highly sensitive and, therefore, accurate at identifying children with DLD, regardless of the presence or absence of a working memory impairment (Archibald & Joanisse, 2009). The researchers also found that combining a nonword repetition task with a sentence repetition task did not more accurately predict DLD than sentence repetition tasks alone when scored at the item level (Archibald & Joanisse, 2009). Yet, Schwob et al.'s (2021) more recent meta-analysis found that scoring a nonword repetition task by counting the number of correct test items, at the whole nonword level, was faster, easier, and had the ability to discriminate between typical developing children and children with DLD.

In Archibald and Joanisse's (2009) study, only 3% of the participants were learning a second language other than English. So, for the purpose of this study, it is essential to examine the effectiveness of sentence repetition tasks in identifying Spanish-speaking and/or bilingual children with DLD. Kapantzoglou et al. (2016) and Armon-Lotem and Meir (2016) found that sentence repetition tasks were a valid predictor

of the grammatical skills of Spanish-speaking and Russian-Hebrew children, respectively. Furthermore, Muñoz and Brimo (2017) and Kapantzoglou et al. (2016) discussed how sentence repetition tasks that are combined with morphology elicitation tasks (i.e. eliciting a child's skills with articles, clitic pronouns, prepositions, subjunctive morphemes, and derivational morphemes) are a valid and reliable way to screen Spanish-speaking children. Armon-Lotem and Meir (2016), in particular, found that nonword and sentence repetition tasks were accurate at identifying DLD in bilingual children regardless of the language tested and that it is essential to determine cut-off scores that are specifically designed for bilingual children through further research and testing, as the cut-off scores used for monolingual children did not provide high diagnostic accuracy for bilingual children.

Kapantzoglou et al. (2016) discussed the need to test more children and set specific cut-off scores to generalize their findings on sentence repetition and morphology elicitation tasks to a larger and more diverse sample. Moreover, Kapantzoglou et al. (2019), found that based on 10% of the tests, inter-rater reliability was estimated to be 94% for the sentence-repetition tasks and 96% for the morphology-elicitation task (Kapantzoglou et al., 2019). This will be examined again during this study. In the end, assessing the predictive power of nonword repetition, sentence repetition, and morphological elicitation measures on the SSLIC is crucial to determine the screener's overall validity and reliability.

Available measures to validate SSLIC. To determine the construct, content, and convergent validity of the SSLIC measure, it is critical that other validated measures be used for the validation. Standardized measures with already established validity can be

used to validate a screening tool (Van der Lely et al., 2011). The Bilingual English-Spanish Assessment (BESA), which will be used in this study, was developed by testing a total of 1,112 children, with and without DLD (Peña et al., 2018). Of those, 420 children were tested in both Spanish and English, 739 children were tested in Spanish, and 632 children were tested in English (Peña et al., 2018). The norms are based on a sample of 756 children and a sample of 198 children who had been previously diagnosed with DLD were used to validate the assessment (Peña et al., 2018). Diagnostic accuracy was assessed during the development of the BESA. The test developers found that the classification accuracy for both the English and Spanish subtests were acceptable regardless of whether the child was a balanced bilingual, bilingual dominant in either Spanish or English, or monolingual in either Spanish or English (Peña et al., 2018). It is important to note that the overall language index score is generated using the highest subtests scores regardless of the language of assessment (Peña et al., 2018). For example, if a student obtains a standard score of 90 on the Spanish morphosyntax and a standard score of 83 on the English morphosyntax, the score of 90 would be used to calculate their overall language index score. This, however, did not affect the BESA's overall classification accuracy and, in some cases, the classification accuracy improved with the use of the best language score (Peña et al., 2018). The BESA has 92% sensitivity and 86% specificity for 4-year-olds, 89% sensitivity and 85% specificity for 5-year-olds, and 96% sensitivity and 92% specificity for 6-year-olds (Peña et al., 2018).

Dynamic assessments have been shown to have high sensitivity and specificity when attempting to identify the presence or absence of DLD in bilingual children (Orellana et al., 2019). In their meta-analysis, Orellana et al. (2019) found two patterns

when using dynamic assessments to evaluate bilingual children with suspected DLD: children with DLD scored lower on both the pre- and posttest portions of the dynamic assessment and were rated as having poorer modifiability, a term used to describe the child's ability to benefit from the teaching portion of the assessment model. In this study, the Dynamic Measure of Oral Narrative Discourse (DYMOND) dynamic assessment protocol will be used, as preliminary research supports the diagnostic accuracy of an English narrative dynamic assessment measure for Spanish-English bilingual children (Petersen et al., 2017). The DYMOND assessment includes a pretest, teaching phase, and a posttest that evaluates a child's ability to tell a narrative and then improve that skill after direct instruction (Frahm, 2021). The DYMOND assessment also provides evaluators with specific procedures that must be used during the teaching phase of the assessment and a rating scale to judge the child's modifiability (Frahm, 2021). Research on the DYMOND assessment revealed similar findings as the aforementioned research on dynamic assessments, i.e. children with DLD scored worse on the posttest and modifiability portions of the DYMOND assessment (Frahm, 2021). Consequently, modifiability ratings on the DYMOND assessment were found to be the best predictor of DLD with 100% sensitivity and 88% specificity after only 1 dynamic assessment session (Clark, 2019; Petersen et al., 2017). A total modifiability index of less than or equal to 10 out of 14 was found to be 90% sensitive and 91% specific (Petersen et al., 2017). The DYMOND is a particularly appropriate assessment tool for this study because the DYMOND assessment was found to be a less-biased method of assessing the presence or absence of DLD in children from culturally and linguistically diverse backgrounds (Frahm, 2021). In the end, an English narrative dynamic assessment, such as the

DYMOND, has been found to have high classification accuracy for bilingual children with and without DLD (Petersen et al., 2017). The researchers, however, believe the DYMOND would not be a suitable screening tool for predominantly Spanish-speaking students as it is fully in English.

Language sample analysis is an appropriate tool for assessing children's language, especially in those who are bilingual (Ebert & Pham, 2017; Heilmann, Miller, & Nockerts, 2010; Kapantzoglou, Fergadiotis, & Restrepo, 2017; Restrepo, 1998). As discussed above morphological markers of DLD can be identified through cloze tasks or language sampling. For the purpose of the current project we will be examining morphology and grammar through cloze tasks and sentence repetition for validation of the SSLIC.

We propose the following questions:

1. Is there preliminary construct, content, and convergent validity evidence for the SSLIC as a screening tool for DLD as measured by:
    a) Correlations with the Bilingual English-Spanish Assessment (BESA) expressive, receptive, sentence repetition, and cloze task raw scores?
    b) Correlations with the Dynamic Measure of Oral Narrative Discourse (DYMOND)?
    c) Differences in group means based on BESA classification?

2. Does the SSLIC show interrater reliability when evaluating for DLD in bilingual and predominantly Spanish-speaking children as assessed through the use of double scoring?

Methods

Participants

The study was approved by the ASU IRB board. Students were recruited by distributing flyers to a local charter school district. We recruited 41 students; of those, fourteen students met criteria after consulting with teachers and the Speech-Language Pathologist and the criteria described below. Parent permission was obtained before starting any testing by means of recruitment packets and consent forms sent home by teachers.

Participant Selection Criteria

All children met the following inclusionary criteria:

- Were between 5-7 years of age
- Passed their school hearing screening.
- Were considered a balanced bilingual or predominantly Spanish-speaking as judged by one of the following criteria:
    - a) Parental reports on an adapted version of the Bilingual Input-Output Survey (BIOS).
    - b) Scored the same or higher on the Spanish SELPS than the English SELPS.
    - c) Scored the same or higher on at least one subtest on the Spanish BESA compared to the English BESA.

Inclusion criteria for the children at risk of DLD:

- Children must score below the cut off standard scores (ex. 86 for 5 year-old and 81 for 6 year-old children) for the BESA best language index.

Inclusionary criteria for the typically developing (TD) children are as follows:

- Children who score equal or ≥ the cut off score for the BESA based on the best language index score for their age, i.e. 86 for 5-year-olds and 81 for 6-year-olds.

Table 1
Demographic Characteristics of Participants

Variable	n	%	Mean	SD
Gender				
Female	5	36%		
Male	9	64%		
Classroom Type				
Dual-language	9	64%		
English-only	5	36%		
Diagnosis				
Typical Language	11	79%		
Developmental Language Disorder	3	21%		
Age at testing (years)			68.28 months	6.498 months
5	10	71%		

	6	3	21%
	7	1	7%
Grade			
Kindergarten	13	93%	
1st Grade	1	7%	

Notes. Values rounded to the nearest whole number and thus one total percentage equates to 99%.

Measures

BIOS

The Bilingual Input-Output Survey, which is a tool that was developed alongside the BESA, was utilized to make judgements about each participants' exposure and use of both Spanish and English (Peña et al., 2018). As testing progressed, the BIOS was adapted to make it easier for the participants' parents to complete. The adapted version asked parents to estimate how often the participant used English, Spanish, and other languages. It also provided parents with an area to describe any speech and/or language concerns they had for their child. This tool was used to determine if a participant had the necessary exposure to and use of Spanish to participate in this study.

SELPS

The Spanish-English Language Proficiency Scale (SELPS) was used to evaluate each participants' proficiency in both Spanish and English. The SELPS is a narrative retell task that uses a wordless picture book to evaluate the participants' abilities with

production, grammaticality, fluency, and vocabulary in both Spanish and English (Smyk et al., 2013). It was also used to determine if a participant had the necessary proficiency in Spanish to participate in this study. It uses a rating scale of 1-4 points for the following areas: vocabulary, grammaticality, sentence length and complexity, and fluency. Smyk et al. (2013) found that the measure was valid for English proficiency assessment.

BESA

The Bilingual English Spanish Assessment (BESA) is a standardized language assessment that tests a participants language skills in both Spanish and English (Peña et al., 2018). It evaluates both morphosyntactic and semantic skills in both languages. Cloze, sentence repetition, and expressive and receptive semantic language tasks are all included on the BESA in both languages. The participant still receives credit for a response even if it is not in the language that the question is asked in as long as it still fits with the prompt. The BESA was used to classify participants as either typical developing or DLD and was compared to the results of the SSLIC to establish validity. The best language index standard score was used to classify participants and correlations were run between the BESA raw scores and SSLIC raw scores.

DYMOND

The Dynamic Measure of Oral Narrative Discourse (DYMOND) is a standardized dynamic assessment that utilizes the test-teach-retest format to assess a child's ability to learn narrative language skills (Petersen et al., 2017). A story is told, the child is asked to repeat the story, visual supports and direct instruction are used to teach the various components of a narrative, a new story is told, and the child is asked to repeat that final story. The DYMOND was used to classify participants as either typical developing or

DLD and was compared to the results of the SSLIC to establish validity. The DYMOND was only used as a comparison measure for students who are balanced bilinguals, as it is entirely in English. This, therefore, makes it inappropriate for monolingual Spanish speakers. However, given this study's small sample size, we cannot definitively prove this.

The DYMOND is scored based on whether the child's story retell included various aspects of the story such as the character, problem, attempt, consequence, etc. Modifiability ratings were also given. These ratings evaluate a child's ability to learn based on their response to prompts, degree of transfer, level of frustration, etc. and provide clinicians with an opportunity to judge the child's overall potential to learn narrative language. These modifiability ratings were found to be the best predictor of DLD (Clark, 2019; Petersen et al., 2017).

SSLIC

The Spanish Screener for Language Impairment in Children is the focus of this study and is designed to screen a child's language skills in Spanish. It is used to determine whether a child is typically developing or if they are at risk for DLD and require further assessment. Cloze and repetition tasks are included on the SSLIC: 27 morphosyntax (5 clitic items, 7 preposition items, 4 derivational morpheme items, 6 subjunctive items, and 5 article items), 5 nonword repetition items, and 10 sentence repetition items were included, making the total possible raw score 42. Scores on the SSLIC will be compared to the BESA and DYMOND (when appropriate) to establish validity.

Procedure

Data Collection

Following recruitment and parent permission, the research team went to a local charter school to administer the SELPS. Then, the Bilingual English-Spanish Assessment (BESA) and DYMOND were administered. Finally, the SSLIC was administered.

The administration of assessments was conducted by graduate and undergraduate student clinicians in ASU's communication disorders program under the supervision of a certified and licensed bilingual speech-language pathologist. Assessments were administered in a random order in a quiet place at the school where the study's participants attend. Audio recordings of the assessment sessions were taken and saved on a secured internet database to review for scoring purposes.

Throughout all testing sessions, breaks were given as needed. At the end of each testing session, the RA gave participants stickers as incentives. When testing was completed, each participant was presented with an age-appropriate bilingual book. Each classroom teacher involved in the study received a $15.00 dollar gift card. Funding for these participant incentives was provided by the JumpStart Grant via the Graduate and Professional Student Association as well as the PRIDE Grant via the College of Health Solutions.

Analyses

RQ 1: Is there preliminary validity evidence for the SSLIC as a screening tool for language impairment? To answer question 1, participants were categorized into two groups using the BESA: TD and DLD. This was based on the aforementioned inclusion criteria. Correlations were run between student SSLIC scores, BESA scores, and

DYMOND scores. Table 2 contains the Pearson correlations for the SSLIC and BESA scores. In addition, an Analysis of variance was run with three different measures of the SSLIC - total score, morphology, sentence repetition, and nonword repetition.

Table 2

Correlations For SSLIC and BESA Spanish Raw Scores

Variab-les	BESA Expres-sive	BESA Recept-ive	BESA SR	BESA Cloze	SSLIC-T	SSLIC-M	SSLIC-SR	SSLIC-NW
BESA Expres-sive		.58*	.68**	.64*	.60*	.49	.50	.38
BESA Recept-ive			.48	.47	.48	.47	.33	.46
BESA SR				.80**	.81**	.78**	.52	.59*
BESA Cloze					.96**	.94**	.71**	.71**
SSLIC-T						.995**	.68**	.77**

SSLIC-M			.62*	.76**
SSLIC-SR				.27
SSLIC-NW				

Notes. BESA SR=BESA Sentence Repetition, BESA Cloze=BESA Cloze Task, SSLIC-T=SSLIC Total score, SSLIC-M=SSLIC Morphosyntax, SSLIC-SR=SSLIC Sentence Repetition, SSLIC-NW=SSLIC Nonword Repetition

* $p < .05$.

** $p < .001$.

There were very strong (above .8), significant (at the .01 level), positive correlations between SSLIC total scores and BESA Spanish expressive raw scores, BESA Spanish sentence repetition raw scores, and BESA Spanish cloze raw scores. There were also very strong (above .8), significant (at the .01), positive correlations between SSLIC morphosyntax scores and BESA Spanish sentence repetition raw scores and BESA Spanish cloze raw scores. Finally, there were very strong (above .8), significant (at the .01 level), positive correlations between SSLIC sentence repetition and nonword repetition scores and BESA Spanish cloze raw scores. There were strong (.6-.79), significant (at the .05 level), positive correlations between the SSLIC total scores and BESA Spanish expressive raw scores. There were also strong (.6-.79), significant (at the .05 level), positive correlations between the SSLIC nonword repetition scores and BESA

Spanish sentence repetition raw scores. Given the small sample size, some of the relations were large but not significant. Table 3 contains the Pearson correlations for the SSLIC and DYMOND scores:

Table 3

Correlations For SSLIC and DYMOND Scores For 11 Participants

Variables	DYMON-D Total Modifiab-ility Score	DYMON-D Judgeme-nt Score	SSLIC-T	SSLIC-M	SSLIC-SR	SSLIC-NW
DYMON-D Total Modifiab-ility Score		.83**	.08	.06	-.30	.37
DYMON-D Judgeme-nt Score			-.08	-.09	-.31	.19
SSLIC-T				.995**	.68**	.77**
SSLIC-M					.62*	.76**
SSLIC-SR						.27
SSLIC-NW						

Notes. SSLIC-T=SSLIC Total score, SSLIC-M=SSLIC Morphosyntax, SSLIC-SR=SSLIC Sentence Repetition, SSLIC-NW=SSLIC Nonword Repetition

* $p < .05$.

** $p < .001$.

There were no significant correlations between SSLIC scores and DYMOND scores.

Table 4

Means, Standard Deviations, and One-Way Analyses of Variance in SSLIC

Measure		TD n = 11		DLD n = 3		$F(1, 12)$^	η2	*p*
		M	*SD*	*M*	*SD*			
SSLIC	Total Potential Raw Score							
Total Score	44	16.1	10.0	5.0	2.0	3.44	.24	.09

Morp-hology	27	11.6	7.8	2.7	2.1	3.58	.23	.08
Sente-nce Rep.	10	0.9	1.4	0.0	0.0	1.12	.09	.31
Nonw-ord Rep.	7	3.2	1.2	2.3	0.6	1.34	.11	.27

Notes. ^$F(1, 12)$ is the case for all measures except SSLIC Total and SSLIC Nonword are $F(1, 11)$ due to missing data for 1 student on these tasks.

Rep.= Repetition

$\eta 2 = 0.01$ indicates a small effect; $\eta 2 = 0.06$ indicates a medium effect; $\eta 2 = 0.14$ indicates a large effect; * $p < .05$. ; ** $p < .001$.

Group mean differences (TD versus DLD) were evaluated using a one-way analysis of variance (ANOVA) on the SSLIC measure. There were no significant mean differences for SSLIC scores as shown below in Table 4. However, we can see graphically that the means were different for SSLIC Total scores with the TD group (0) having a mean total score of 16 and the DLD group (1) a total score of 5 (shown in Figures A and B) as well

as SSLIC Morphosynax scores with the TD group having a mean of 11.55 and the DLD group having a mean of 2.67 (shown in Figures A and B). Further, the effect sizes between groups are large effect sizes on the SSLIC total score and morphology measures and medium on the sentence and nonword repetition measures supporting differences between groups. See table 4 and figures A and B.

Figure A

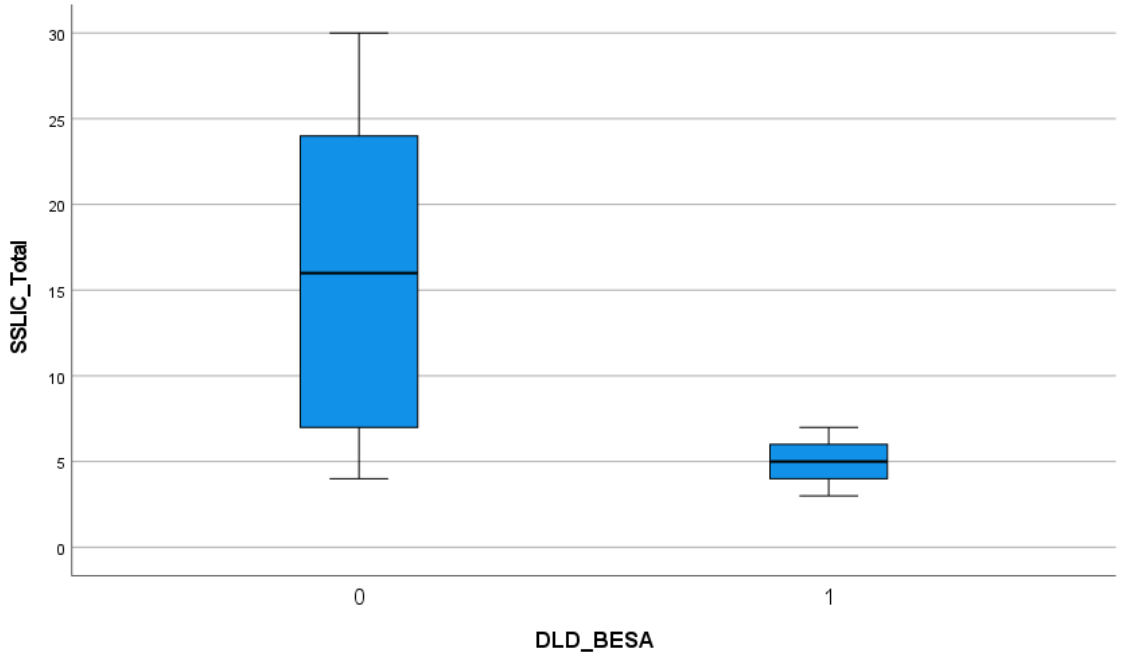
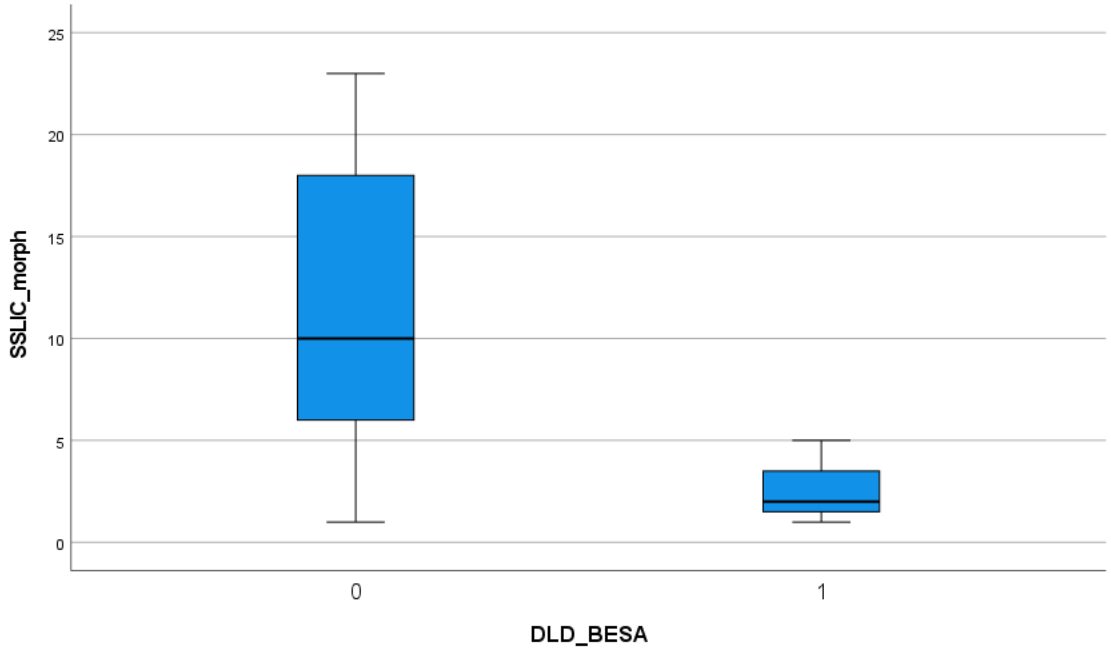


Figure B



Notes. 0 = Typical Developing (TD), 1 = Developmental Language Disorder (DLD).

RQ 2: Does the SSLIC show interrater reliability when evaluating for DLD in bilingual and predominantly Spanish-speaking children as assessed through the use of double scoring?

To answer this question, 20% of the assessments (3) were double scored to determine interrater reliability. Point to point analysis was completed to compare the results of two raters for three assessments. The mean percent agreement for the morphology elicitation task was 95%. The mean percent agreement for the nonword repetition task was 95%. This task was scored at the item level, i.e. either correct or incorrect for each nonword production depending on whether the whole nonword was repeated accurately or inaccurately. The mean percent agreement for the sentence

repetition task was 100%. These values show excellent interrater reliability. See Table 5 for scores given by each rater and percent agreement per student per subtest.

Table 5

Interrater Agreement on the SSLIC based on Point to Point Analysis

Student ID	SSLIC Subtest	Total Items	Percent Agreement
6	Morphology Elicitation	27	.93
	Nonword Repetition	7	1
	Sentence Repetition	10	1
31	Morphology Elicitation	27	1
	Nonword Repetition	7	.86
	Sentence Repetition	10	1
38	Morphology Elicitation	27	.93
	Nonword Repetition	7	1
	Sentence Repetition	10	1

Discussion

The results of the study answered the first question addressed: what is the ability of the SSLIC measure to differentiate between TD and DLD groups based on correlations with full measures, i.e. the BESA and DYMOND, and group mean differences? The results of this preliminary study indicate that the SSLIC measure correlates with the BESA which is a validated measure for the bilingual population. Further, the measure differentiates between TD and DLD groups based on preliminary results. Although some of the differences were not significant, the effect sizes indicate that there are group differences on the SSLIC total, sentence repetition, nonword repetition and morphology. Therefore, this study supports that the SSLIC has the potential to be used to screen the language abilities of balanced Spanish-English bilingual kindergarten and 1st grade students; however, further validation is needed given the small sample size. Despite this study's small sample size, many significant correlations were found between each of the tasks on the SSLIC and the BESA. This supports that morphology elicitation, nonword repetition, and sentence repetition can be used to identify Spanish-speaking children with DLD. These results are consistent with those of studies that use similar measures that have been found to differentiate monolingual and bilingual DLD children (Armon-Lotem and Meir, 2016; Castilla et al., 2016; Castilla et al., 2021; Estes et al., 2007; Girbau & Schwartz, 2007; Kapantzoglou et al., 2016; Morgan et al., 2013; Muñoz & Brimo, 2017; Ortiz, 2021; Schwob et al., 2021).

The second process for validation was the examination of correlations between two measures that differentiate TD and DLD. Correlations between the BESA and SSLIC indicate that the tasks included on the SSLIC, i.e. morphology elicitation, nonword

repetition, and sentence repetition support the SSLIC's content validity, as children's scores on the SSLIC had both strong and very strong correlations with children's scores on the BESA. The SSLIC was originally compared to the CELF-4S; however, this test was found to over-identify children as having DLD due to the fact it is modeled after its English counterpart and does not account for children's various levels of bilingualism (Barragan et al, 2018; Restrepo et al., 2010). The BESA allows evaluators to assess a child's expressive and receptive language skills in both Spanish and English, which accounts for each child's varying levels of bilingualism and provides more accurate ratings of the child's skills (Peña et al., 2018). Ultimately, the BESA has been found to be accurate at identifying DLD within Spanish-English bilingual children (Peña et al., 2018). So, correlations with the BESA suggest that the SSLIC is accurate at identifying DLD.

The results of the SSLIC did not correlate with the results of the DYMOND and no significant mean differences were found between the TD group and the DLD group. This would suggest that the SSLIC is not accurate at identifying DLD in the predominantly Spanish-speaking population. Yet, the researchers believe that the DYMOND was ultimately not a suitable assessment for the participants of this study. The DYMOND was originally selected based on its ability to assess bilingual Spanish-English children. However, this study's participants were predominantly Spanish-speaking students and the DYMOND is an English assessment. Additionally, only 1 student with DLD in this study was able to complete the DYMOND. So, the sample size was not large enough to accurately assess for correlations between these two measures nor group mean differences. The discrepancy with the language of test

administration and small sample size likely accounts for the lack of correlations found between these two measures.

The second research question examined the interrater reliability of the SSLIC. The SSLIC was found to have excellent interrater reliability based on point to point analysis. These results support those of Kapantzoglou et al. (2016) who found excellent reliability on the SSLIC's morphology elicitation and sentence repetition tasks.

These results serve as preliminary evidence that the SSLIC is both a valid and reliable tool for the screening of DLD in Spanish-speaking kindergarten and 1st grade students. At this time, the researchers believe that the SSLIC is most appropriate for primarily Spanish-speaking students. The continued validation of the SSLIC will help remediate the need for valid and reliable screening tools within the field of speech-language pathology (Molanda and Oetting, 2021). It will also fill a great need, as there are currently no Spanish screening tools available commercially to practicing speech-language pathologists. Lastly, a quality screening tool for DLD with Spanish-speaking children will help to reduce the identification and service delivery discrepancies that currently exist in US public schools (Morgan et al., 2017; Yamasaki and Luk, 2018).

Limitations

The small sample size in this study made the statistical analysis for predicting group membership inappropriate and results indicate that it was underpowered. Additionally, it's hypothesized that a small sample size likely accounts for the moderate reliability and lack of mean differences found on the DYMOND assessment. Many of the

students that were originally tested were found to be English-dominant bilingual students. So, they were not included in this study. This reinforces the fact that the SSLIC is designed for primarily Spanish-speaking students. Further, clinical observations of participants' performance suggest that a child needs to have high levels of Spanish input and output to be considered an appropriate recipient of the SSLIC. Further, clinical observations also suggest that a child needs to have high levels of English input and output to be considered an appropriate recipient of the DYMOND.

The DYMOND assessment, although it is appropriate for balanced bilingual students, was not administered to many of the Spanish-dominant students because it is an English-based assessment and is therefore not appropriate for this population. The DYMOND assessment was also originally designed for 1st grade students. So, it may not accurately represent the language abilities of some of the participants of this study.

Consistency with the administration of the SSLIC can also be improved upon. The researchers used the demonstration prompt from the BESA when administering and providing instruction for practice items for the SSLIC cloze tasks (i.e. fíjate en como yo lo digo). Moreover, the sentence repetition task did not include any specific prompts. So, the administrator's personal clinical experience was used to decide how instruction should be given and if repetitions of stimuli items should be given (ex. for times when the participant was not paying attention and/or became distracted).

Future Directions

To further validate the SSLIC additional children need to be recruited and discriminant analyses can be used to classify participants into separate groups and

examine the accuracy (Anaya, Peña, & Bedore, 2018; Kapantzoglou, Fergadiotis, & Restrepo, 2017; Lazewnick, 2019). However, we were not able to recruit the necessary sample size after a year.A logistic regression should be conducted in future studies to determine if the SSLIC results predict group members (i.e. TD and DLD). A discriminant analysis should also be conducted in future studies to determine the appropriate cut-off scores for passing versus referral in children with different levels of language proficiency.

Further validation is required for primarily Spanish-speaking, monolingual Spanish, and Spanish-English balanced bilinguals. A comparison of the performance and, possibly, a determination of appropriate cut score for each group is needed. Comparing SSLIC performance of students from Spanish-English dual-language classrooms and English-only classrooms would also be beneficial. Finally, a comparison involving the amount of instruction on practice items of the SSLIC cloze tasks and repetition tasks should also be conducted. Various levels of direct instruction during demonstration items should be compared to determine the amount of instruction that will best support the SSLIC's accuracy. Direct instruction could help clarify the purpose of the task and what is expected of the student. This should ensure that the SSLIC is not over-identifying children as having DLD. The use of additional nonword and sentence repetition task practice items should also be considered. A specific protocol, record form, training, and script should be considered when considering any of these comparisons. These resources would provide additional support to clinicians and promote standardization of the SSLIC measures.

Considering the difficulties that many of this study's participants had with the DYMOND, the development of a Spanish version of the DYMOND should be considered in future studies.

Clinical Implications

The SSLIC has the potential to be used to screen the language abilities of Spanish-speaking, primarily Spanish-speaking students, and balanced Spanish-English bilingual. This can likely help reduce the service delivery discrepancies that have been found between Spanish-speaking students in US public schools and their English-speaking peers. Further establishing the validity and reliability of the SSLIC on a larger scale would ensure that the SSLIC is accurate and consistent. Then, more students who need further language assessment would receive it while also mitigating the current problem of over-identification within this population.

The DYMOND is not suitable for monolingual Spanish-speaking students. It should only be used for balanced Spanish-English bilingual students or English-dominant bilingual students.

Summary

The purpose of this study was to evaluate the SSLIC's ability to screen Spanish-English bilingual and Spanish-speaking monolingual children for DLD, as there are currently few quality screening tools available that can accurately do this. This is likely a factor that leads to inaccurate identification of DLD and unnecessary delivery of speech and language services to these children. The SSLIC measures oral language skills

in Spanish-speaking children by conducting morphology elicitation, nonword repetition, and sentence repetition tasks. The results on these measures were compared to other validated measures, Significant correlations were found between the SSLIC's scores and various scores on the BESA. No significant mean differences for the SSLIC were found. Yet, the mean total and morphosyntax scores were different between the two groups upon visual inspection and examination of the effect sizes. Double scoring revealed that the SSLIC has high interrater reliability. Overall, despite a small sample size, this study supports that the SSLIC could feasibly be used to screen for DLD in Spanish-speaking kindergarten and 1st grade students and suggests that the SSLIC is both valid and reliable. Continued validation of the SSLIC on a larger scale should be conducted.

REFERENCES

Anaya, J. B., Peña, E. D., & Bedore, L. M. (2018). Conceptual scoring and classification accuracy of vocabulary testing in bilingual children. *Language, Speech & Hearing Services in Schools*, *49*(1), 85–97. https://doi.org/10.1044/2017_LSHSS-16-0081

Archibald, L. M. D. & 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. https://doi.org/10.1044/1092-4388(2009/08-0099)

Armon-Lotem, S. & Meir, N. (2016). Diagnostic accuracy of repetition tasks for the identification of specific language impairment (SLI) in bilingual children: evidence from Russian and Hebrew. *International Journal of Language & Communication Disorders*, *51*(6), 715–731. https://doi.org/10.1111/1460-6984.12242

Barragan, B., Castilla-Earls, A., Martinez-Nieto, L., Restrepo, M. A., & Gray, S. (2018). Performance of low-income dual language learners attending english-only schools on the clinical evaluation of language fundamentals–fourth edition, Spanish. *Language, Speech & Hearing Services in Schools*, *49*(2), 292–305. https://doi.org/10.1044/2017_LSHSS-17-0013

Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, *52*(6), 671–680. https://doi.org/10.1111/1460-6984.12335

Castilla-Earls A., Bedore, L., Rojas, R., Fabiano-Smith, L., Pruitt-Lord, S., Restrepo, M. A., & Peña, E. (2020). Beyond scores: Using converging evidence to determine speech and language services eligibility for dual language learners. *American Journal of Speech-Language Pathology*, *29*(3), 1116–1132. https://doi.org/10.1044/2020_AJSLP-19-00179

Castilla-Earls, A., Pérez-Leroux, A. T., Fulcher-Rood, K., & Barr, C. (2021). Morphological errors in spanish-speaking bilingual children with and without developmental language disorders. *Language, Speech & Hearing Services in Schools*, *52*(2), 497–511. https://doi.org/10.1044/2020_LSHSS-20-00017

Castilla-Earls A. P., Restrepo, M. A., Pérez-Leroux, A. T., Gray, S., Holmes, P., Gail, D., & Chen, Z. (2016). Interactions between bilingual effects and language impairment: Exploring grammatical markers in Spanish-speaking bilingual children. *Applied Psycholinguistics*, *37*(5), 1147–1173. https://doi.org/10.1017/S0142716415000521

Clark, K. D. (2019). *The Cross-Validation of the Classification Accuracy of a Dynamic Assessment of Narrative Language for School-Age Children with and Without Language Disorder*. ProQuest Dissertations Publishing.

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. https://doi.org/10.1044/1092-4388(2010/10-0093)

Ebert, K. D. & Pham, G. (2017). Synthesizing information from language samples and standardized tests in school-age bilingual assessment. *Language, Speech & Hearing Services in Schools*, *48*(1), 42–55. https://doi.org/10.1044/2016_LSHSS-16-0007

Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the Nonword Repetition Performance of Children With and Without Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Frahm, A. E. (2021). *Norming a Dynamic Assessment of Narrative Language for Diverse School-Age Children with and without Language Disorder: A Preliminary Psychometric Study*. ProQuest Dissertations Publishing.

Girbau D. & Schwartz, R. G. (2007). Non-word repetition in Spanish-speaking children with Specific Language Impairment (SLI). *International Journal of Language & Communication Disorders*, *42*(1), 59–75. https://doi.org/10.1080/13682820600783210

Heilmann, J. J., Miller, J. F., & Nockerts, A. (2010). Using Language Sample Databases. *Language, Speech & Hearing Services in Schools*, *41*(1), 84–95. https://doi.org/10.1044/0161-1461(2009/08-0075)

Kapantzoglou, M., Fergadiotis, G., & Restrepo, M. A. (2017). Language sample analysis and elicitation technique effects in bilingual children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, *60*(10), 2852–2864. https://doi.org/10.1044/2017_JSLHR-L-16-0335

Kapantzoglou, M., Thompson, M. S., Gray, S., & Restrepo, M. A. (2016). Assessing measurement invariance for spanish sentence repetition and morphology elicitation tasks. *Journal of Speech, Language, and Hearing Research*, *59*(2), 254–266. https://doi.org/10.1044/2015_JSLHR-L-14-0319

Karem, R. W. & Washington, K. N. (2021). The cultural and diagnostic appropriateness of standardized assessments for dual language learners: A focus on jamaican

preschoolers. *Language, Speech & Hearing Services in Schools*, *52*(3), 807–826. https://doi.org/10.1044/2021_LSHSS-20-00106

Lazewnik R., Creaghead, N. A., Smith, A. B., Prendeville, J.-A., Raisor-Becker, L., & Silbert, N. (2019). Identifiers of language impairment for spanish-english dual language learners. *Language, Speech & Hearing Services in Schools*, *50*(1), 126–137. https://doi.org/10.1044/2018_LSHSS-17-0046

Leaders Project. (2013, October 29). *Test review: PLS-5 Spanish*. Leaders Project. Retrieved August 22, 2022, from https://www.leadersproject.org/2013/10/29/test-review-pls-5-spanish/

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. https://doi.org/10.1044/2015_AJSLP-14-0061

Moland C. W. & Oetting, J. B. (2021). Comparison of the diagnostic evaluation of language variation–screening test risk subtest to two other screeners for low-income prekindergartners who speak african american english and live in the urban south. *American Journal of Speech-Language Pathology*, *30*(6), 2528–2541. https://doi.org/10.1044/2021_AJSLP-20-00270

Morgan, G. P., Restrepo, M. A., & Auza, A. (2013). Comparison of Spanish morphology in monolingual and Spanish–English bilingual children with and without language impairment. *Bilingualism (Cambridge, England)*, *16*(3), 578–596. https://doi.org/10.1017/S1366728912000697

Morgan, P. L., Farkas, G., Hillemeier, M. M., Li, H., Pun, W. H., & Cook, M. (2017). Cross-Cohort Evidence of Disparities in Service Receipt for Speech or Language Impairments. *Exceptional Children*, *84*(1), 27–41. https://doi.org/10.1177/0014402917718341

Muñoz, M. L. & Brimo, D. (2017). Evidence for the construct validity of two grammatical tasks to screen for language impairment in Spanish speaking children. *Evidence Based Communication Assessment and Intervention*, *11*(3-4), 146–150. https://doi.org/10.1080/17489539.2017.1339480

National Center for Education Statistics. (2020). *English language learner (ELL) students enrolled in public elementary and secondary schools, by home language, grade, and selected student characteristics: Selected years, 2008-09 through fall 2018* [Table]. nces.ed.gov. https://nces.ed.gov/programs/digest/d20/tables/dt20_204.27.asp

National Center for Education Statistics. (2022). Racial/Ethnic Enrollment in Public Schools. *Condition of Education*. U.S. Department of Education, Institute of

Education Sciences. Retrieved [December 15, 2022], from https://nces.ed.gov/programs/coe/indicator/cge.

Orellana, C. I., Wada, R., & Gillam, R. B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. *American Journal of Speech-Language Pathology, 28*(3), 1298–1317. https://doi.org/10.1044/2019_AJSLP-18-0202

Ortiz, J. A. (2021). Using nonword repetition to identify language impairment in bilingual children: A meta-analysis of diagnostic accuracy. *American Journal of Speech-Language Pathology*, *30*(5), 2275–2295. https://doi.org/10.1044/2021_AJSLP-20-00237

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. https://doi.org/10.1044/1092-4388(2012/12-0050)

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2018). *Bilingual English-Spanish Assessment Manual*. Paul H Brookes Publishing Co.

Petersen, D. B., Chanthongthip, H., Ukrainetz, T. A., Spencer, T. D., & Steeve, R. W. (2017). Dynamic assessment of narratives: Efficient, accurate identification of language impairment in bilingual students. *Journal of Speech, Language, and Hearing Research*, *60*(4), 983–998. https://doi.org/10.1044/2016_JSLHR-L-15-0426

Pezold, M. J., Imgrund, C. M., & Storkel, H. L. (2020). Using computer programs for language sample analysis. *Language, Speech & Hearing Services in Schools*, *51*(1), 103–114. https://doi.org/10.1044/2019_LSHSS-18-0148

Plante, E. & Vance, R. (1994). Selection of Preschool Language Tests: A Data-Based Approach. *Language, Speech & Hearing Services in Schools*, *25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Restrepo M. A., Gorin, J. S., Gray, S., Morgan, G. P., & Barona, N. (2010). *Development of a Language Impairment Screener for Spanish Speaking Children : SSLIC: Phase 1--Task Development*. Distributed by ERIC Clearinghouse.

Restrepo, M. A. (1998). Identifiers of Predominantly Spanish-Speaking Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*, *41*(6), 1398–1411. https://doi.org/10.1044/jslhr.4106.1398

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using Nonword Repetition to Identify Developmental

Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *64*(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Smyk, E., Restrepo, M. A., Gorin, J. S., & Gray, S. (2013). Development and validation of the Spanish-English language proficiency scale (SELPS). *Language, Speech & Hearing Services in Schools*, *44*(3), 252–265. https://doi.org/10.1044/0161-1461(2013/12-0074)

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International journal of medical education, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Tomblin, B., Records, N., Buckwalter, P., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40* (6), 1245-1260. https://doi.org/10.1044/jslhr.4006.1245

United States Census Bureau (2021). United States. data.census.gov. [https://data.census.gov/profile?q=United+States&g=0100000US]

Van Der Lely, H. K. J., Payne, E., & McClelland, A. (2011). An investigation to validate the grammar and phonology screening (GAPS) test to identify children with specific language impairment. *PloS One*, *6*(7), e22432–e22432. https://doi.org/10.1371/journal.pone.0022432

Yamasaki, B. L. & Luk, G. (2018). Eligibility for special education in elementary school: The role of diverse language experiences. *Language, Speech & Hearing Services in Schools*, *49*(4), 889–901. https://doi.org/10.1044/2018_LSHSS-DYSLC-18-0006