2021-06-17

The Use of Nonword Repetition Tasks in the Assessment of Developmental Language Disorder in Bilingual Children

Kirsten Kelly
Brigham Young University

The Use of Nonword Repetition Tasks in the Assessment of Developmental Language Disorder in Bilingual Children

Kirsten Kelly

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Connie Summers, Chair
Kathryn Cabbage
Erika Feinauer

Department of Communication Disorders

Brigham Young University

ABSTRACT

The Use of Nonword Repetition Tasks in the Assessment of
Developmental Language Disorder in
Bilingual Children

Kirsten Kelly
Department of Communication Disorders, BYU
Master of Science

To address the needs of the growing number of Spanish-English bilingual children in the United States, Nonword Repetition (NWR) tasks were created to reduce testing bias in the assessment and diagnosis of children with developmental language disorder (DLD). Several studies have shown promising results in the use of NWR tasks; however, fewer studies have addressed questions such as the use of different scoring methods or analyzing error patterns. Thus, this study was conducted to address these gaps in the research. An English and a Spanish NWR task were administered to 26 Spanish-English bilingual school aged children (6;0-9;4). Two different scoring methods (percent phoneme correct and whole word scoring) were compared for diagnostic accuracy and the types and frequency of errors were analyzed. Both scoring methods showed statistically significant differences between groups (participants with DLD and those with typically developing language). Whole word scoring in Spanish had the best diagnostic accuracy, according to sensitivity, specificity, and likelihood ratio measures. However, due to the small number of nonwords that any participant repeated correctly, this may not be a clinically practical scoring method. The Spanish NWR task was a better measure than the English NWR task in identifying children with DLD, suggesting that Spanish NWR could be used to assess DLD in bilingual children. Participants with DLD produced more consonant, vowel, substitution, and omission errors than those with typically developing language. There was no difference between groups for addition errors. Significantly more omission errors were made in Spanish, likely due to the longer nonwords. The longer nonwords may be key in distinguishing between typically developing children and those with DLD. These results have the potential to inform future clinical practices in selecting, scoring, and analyzing NWR tasks.

Keywords: nonword repetition tasks, bilingual, developmental language disorder, percent phoneme correct, whole word, errors

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DESCRIPTION OF THESIS STRUCTURE AND CONTENT

This thesis, *Using Nonword Repetition Tasks to Assess for Developmental Language Disorder in Bilingual Children*, is written in a hybrid format, bringing together aspects of the traditional thesis requirements and journal publication formats. The beginning pages of this thesis reflect requirements for submission to the university. The thesis itself is presented as a journal article and conforms to style requirements for submitting research reports to education journals. The annotated bibliography is in Appendix A, Appendix B includes an Institutional Review Board statement, and the study's instruments are in Appendix C.

Introduction

Over the past 30 years the number of Spanish-English bilingual children in public schools in the United States (US) has risen dramatically. Latino children are now the largest minority group in the United States and in 2018 they accounted for 25% of children in the US (Federal Interagency Forum on Child and Family Statistics, 2020). Among children learning English as a second language, Spanish-speaking children are the largest group and the numbers are expected to continue growing. Yet, Latino children are more likely to be misidentified for services and experience academic delays compared to other minority students learning English even when they come from a similar socioeconomic background (Guiberson & Rodriguez, 2013; Gutiérrez-Clellan & Simon-Cereijido, 2010). With the rising number of bilingual children in schools, speech-language pathologists must use effective and efficient methods to assess these children to determine any speech or language difficulties they may have, including developmental language disorder (DLD). Developmental language disorder is a language disorder not associated with a known biological cause (Paul et al., 2018) and will be the focus of this paper.

Traditional Testing

Traditional norm referenced tests are commonly used to diagnose language disorders but may not be adequate to use with bilingual children. Traditional norm referenced tests are inherently biased against minority students. These tests place emphasis on experiential history and vocabulary knowledge that minority children may not share with monolingual English-speaking students. A child's experience with a given language may determine their performance on these measures (Campbell et al., 1997; Dollaghan & Campbell, 1998; Kohnert et al., 2006). Because bilingual children's language knowledge is spread across two languages, testing in just one language will show only a partial snapshot of their language knowledge and

experience. Children may have certain words in their vocabulary for one language, but not the other, depending on the use of those words in each language. Most norm referenced tests are administered in English and normed for monolingual English speakers. Thus, they do not account for language differences in bilingual students. Due to this bias, many bilingual children are misidentified and Spanish-speaking English language learners are often under or overrepresented in special education. They may not be receiving the services they need (Gutiérrez-Clellan & Simon-Cereijido, 2010).

Nonword Repetition Tasks

Nonword repetition (NWR) tasks were created to help reduce bias and provide a more efficient and effective way to help diagnose culturally and linguistically diverse children, including bilingual children, with developmental language disorder, when used with other assessment measures. A nonword repetition task is a working memory task that includes nonsense words that adhere to the phonotactic rules of the language the test is intended to mimic. Children listen to a nonword and then repeat it. The nonwords have appropriate sound combinations, but no semantic value (Guiberson & Rodriguez, 2013; Kohnert et al., 2006). As such, NWR tasks are processing-dependent measures, not knowledge-dependent measures. Thus, NWR tasks test fundamental language processing skills and do not rely on prior language knowledge like traditional tests (Dollaghan & Campbell, 1998), helping reduce test bias for children from diverse backgrounds. Roy and Chiat (2004) reported that NWR tasks are culturally unbiased and independent of IQ. They give diagnostically meaningful information for children regardless of their socioeconomic, ethnic, or educational background (Dollaghan & Campbell, 1998). For example, in Dollaghan and Campbell's (1998) study they found that African-American children's scores on an NWR task did not differ significantly from the

scores of Caucasian children. However, in other traditional norm referenced tests, there is often a discrepancy between their scores due to inherent testing bias.

Language Correlations With NWR Tasks

Because NWR is a language-processing measure, children's scores on an NWR task provide information about their language learning abilities. Performance on NWR tasks has been correlated with grammar and grammatical complexity in expressive language, larger expressive and receptive vocabularies, and word learning abilities (Adlof & Patten, 2017; Girbau, 2016; Roy & Chiat, 2004; Summers et al., 2010). Many of the skills necessary for learning new words and grammatical forms are also necessary to complete a nonword repetition task. Completing a nonword repetition task requires a child to store the phonological segments in their working memory, or phonological loop (Baddeley, 1986). Researchers have described this process using varying terms including working memory, verbal working memory, and phonological short-term memory. For the purpose of this paper, the term "working memory" will be used. In an NWR task, a child must listen to the nonword, break it down into its individual phonemes while keeping the order stored in working memory, and then organize the output to recite it back. They are both storing and processing information for a short amount of time. If there are deficits in any of these steps, it could have a negative influence on language learning tasks (Dollaghan & Campbell, 1998; Gathercole, 2006; Gutiérrez-Clellan & Simon-Cereijido, 2010).

In learning new words, a child must be able to hear novel stimuli and immediately repeat it, just like in a nonword repetition task (Guiberson & Rodriguez, 2013). Being able to store information in their working memory allows them to process and learn new words. Working memory is also important to store grammatical forms in the learning process (Girbau, 2016; Roy & Chiat, 2004; Summers et al., 2010). Researchers have found that children with learning

disabilities, attention deficits, reading disabilities, and language impairments may struggle with working memory (Gutiérrez-Clellan & Simon-Cereijido, 2010). Due to these connections, NWR tasks may be useful in determining the language learning abilities of bilingual students, instead of focusing on their current language experience.

Language Experience and NWR Tasks

The nonwords in an NWR task should be equally unfamiliar to all children to ensure that the task does not favor children with more language experience (Dollaghan & Campbell, 1998). When hearing new words, children with more language experience can rely more on their stored knowledge to create connections to the new word. Children with less language experience must rely more on their working memory (Summers et al., 2010). Roy and Chiat (2004) administered repetition tasks with real words and nonwords to children to determine how familiarity to words affected the scores. They found that younger children, who had less word familiarity, performed more poorly on these measures than older children who had more experience with words.

To account for this, Dollaghan and Campbell (1998) recognized the need to create a list of nonwords that minimized the familiarity to real words. Nonwords can be judged to be more or less "wordlike." Nonwords are more wordlike if they have syllables that correspond to lexical items or if they have a high predictability of individual phonemes (Dollaghan & Campbell, 1998; Summers et al., 2010). For example, the nonword "glistering" is a high-wordlike nonword, while the nonword "teivak" is low-wordlike (Summers et al., 2010). In a nonword repetition task, the less wordlike the nonsense words are, the less children can rely on previously stored knowledge to help construct the phonological representation (Gathercole, 2006). The less wordlike the nonwords are, the less biased they are for children with less language learning experience. It places children on a level playing field, as they cannot rely on their stored vocabulary knowledge

as much. Thus, using their working memory becomes important, which also predicts their language learning skills.

Nonword Repetition Tasks for Bilingual Children

Using NWR tasks with nonwords that only reflect the phonotactic properties of English with bilingual children does not fully minimize bias for this group. There is bias for bilingual children because they do not have the same phonetic experience with English sounds as monolingual children do. To account for this discrepancy, Spanish nonword repetition lists have been created (Ebert et al., 2008; Gutiérrez-Clellan & Simon-Cereijido, 2010). These lists include the phonology and stress patterns of Spanish. A Spanish nonword repetition task is important for Spanish-English bilingual children because if they have more exposure to Spanish, it may be easier for them to understand the phonological forms that are more similar to Spanish words. Bilingual children have various levels of exposure to each language and the phonological structure of one language may affect performance on a nonword repetition task in another language (Gibson et al., 2014). Exposure to the different languages will change with context and throughout time (Gutiérrez-Clellan & Simon-Cereijido, 2010).

Diagnostic Accuracy

While steps can be taken to prevent bias in nonword repetition tasks, the biggest question needing to be answered is if it is a diagnostically accurate test to use to help distinguish between children with a DLD and those without. Dollaghan and Campbell (1998), Gutiérrez-Clellan and Simon-Cereijido (2010), and Kohnert et al. (2006) found that on a nonword repetition task using the list of nonwords created by Dollaghan and Campbell, the typically developing group scored significantly higher than the language disordered group. Gathercole et al. (1994) reported that in a study done using The Children's Test of Nonword Repetition (CNRep), the scores on the test

perfectly discriminated between children with DLD and those without. Gray (2003) also found that typically developing children scored significantly higher than children with DLD on the CNRep.

While there is compelling evidence for the diagnostic accuracy in English speaking children with and without DLD, bilingual children do not exhibit the exact same patterns in English NWR tasks. As children learn two languages, they have various exposures to each language, so the accuracy of distinguishing them on a nonword repetition task would vary based on the language of the test (Gutiérrez-Clellan & Simon-Cereijido, 2010). Gutiérrez-Clellan and Simon-Cereijido (2010) found that administering a nonword repetition task in only one language with bilingual children was inadequate to diagnose DLD. Some children with typical language were misidentified as having DLD when information from only one test was analyzed. Kohnert et al. (2006) and Windsor et al. (2010) found that typically developing Spanish-English bilingual children's scores overlapped somewhat with the monolingual English children who had a language disorder on an English nonword repetition task with monolingual English and Spanish-English bilingual children. However, Windsor et al. (2010) found that on a Spanish nonword repetition task, the typically developing bilingual children scored better than the other groups. The English nonword repetition task alone may not be sufficient to correctly differentiate bilingual children with and without DLD. It appears that to get the most accurate information about the language processing skills of Spanish-English bilingual children, they should be tested in both Spanish and English. Testing in both languages may be crucial to get a sense of the child's true language abilities. Bilingual children's language skills are variable across each language, so even testing their dominant language does not show their full skill set. This is evidence that testing bilingual children in only one language, even with a task that is not based

on vocabulary knowledge, such as nonword repetition, does not have sufficient diagnostic accuracy (Gutiérrez-Clellan & Simon-Cereijido, 2010).

Measures of Diagnostic Accuracy

To determine the diagnostic accuracy of nonword repetition tasks, sensitivity, specificity and likelihood ratios are commonly calculated. These measures are meant to explore how accurately children with developmental language disorder are identified as such and typical children are identified as such. Sensitivity shows how accurately children with DLD are identified as such by a given test. Specificity shows how accurately children with typical language are identified as such (Plante & Vance, 1994). Studies of NWR have not reported consistent measures of sensitivity and specificity. Several studies have found higher specificity rates (Gray, 2003; Guiberson & Rodriguez, 2013), while others have found higher sensitivity rates (Girbau, 2016; Girbau & Schwartz, 2007).

Even in studies that have included more than one NWR task to represent the languages of the participants, specificity and sensitivity rates have also been inconsistent. One study found low sensitivity for bilingual children on an English and Spanish nonword repetition task and moderate specificity on each test individually, but good specificity when the NWR task scores were combined (Gutiérrez-Clellan & Simon-Cereijido, 2010). However, another study found high sensitivity for bilingual children on the NWR task they used, but only fair specificity (Thordardottir & Brandeker, 2013). A third study found that there was low specificity for bilingual children on an English NWR task, but moderate specificity on a Spanish NWR task (Windsor et al., 2010). Due to these varying rates for bilingual children, more research is needed in this area.

Another measure to determine diagnostic accuracy is likelihood ratios. Likelihood ratios are less susceptible to variations in the base rate of a sample than sensitivity and specificity and may be preferred over rates of sensitivity and specificity (Dollaghan, 2007). Likelihood ratios can range anywhere between 0 to infinity as they indicate the likelihood of having a disorder (positive likelihood) or ruling out a disorder (negative likelihood). Similar to sensitivity and specificity, likelihood ratios of NWR tasks have been inconsistent. Dollaghan and Campbell (1998) found a positive and negative likelihood ratio that indicated strong diagnostic accuracy on their English nonword repetition task with monolingual English speakers. However, another study that included bilingual speakers found likelihood ratios for the same English NWR task that could rule out a language disorder, but were not sufficient to rule in a disorder (Kohnert et al., 2006). Similarly, Spanish NWR tasks have been found to have strong likelihood ratios in some studies (Girbau, 2016; Girbau & Schwartz, 2007) and weaker likelihood ratios in other studies (Guiberson & Rodriguez, 2013).

Gutiérrez-Clellan and Simon-Cereijido (2010) and Windsor et al. (2010) calculated likelihood ratios for both Spanish and English nonword repetition tasks. Gutiérrez-Clellan and Simon-Cereijido (2010) found a positive likelihood ratio that was highly suggestive of a disorder for the two tests combined. Windsor et al. (2010) found that for the English NWR task, the only likelihood ratio that had sufficient diagnostic accuracy was the negative likelihood ratio for the bilingual group. For the Spanish NWR task, none of the likelihood ratios had sufficient diagnostic accuracy on their own. Given the inconsistent findings in the literature, it could be that using English and Spanish NWR scores combined in some format may improve the diagnostic accuracy for bilingual children. Further research is needed in this area.

Scoring

The type of scores used in NWR may affect the diagnostic accuracy. There have been two major ways of scoring used for nonword repetition tasks. One method of scoring is by percent phoneme correct (PPC). The scoring rules set apart by Dollaghan and Campbell (1998) have been commonly used for this scoring. In this level of scoring the number of phonemes that the child said correctly was divided by the total number of phonemes to calculate a percent phoneme correct. Phoneme substitutions and omissions were marked as incorrect. Distortions were counted as correct. Additions were not counted as errors because the main goal was to look at information that was lost. Vowels were used as a syllable anchor so when a child omitted a consonant the researchers scoring the test would line up the child's production to the closest syllable. Several studies have found success with this scoring strategy being able to distinguish children with and without developmental language disorder.

Some studies have used a second strategy for scoring NWR tasks called whole word or item level scoring. For this scoring method the child receives either a 1 or a 0 for each word. They must repeat the whole word correctly to get a 1 and if there are any errors, they receive a 0. Guiberson and Rodriguez (2013) used some of the same scoring rules as for percent phoneme correct scoring, so distortions were still counted as correct and additions were not counted as incorrect. Roy and Chiat (2004) counted additions as incorrect. Whole word scoring has the potential to be clinically useful as it would significantly reduce the amount of time and effort a clinician needs to spend on scoring (Guiberson & Rodriguez, 2013; Roy & Chiat, 2004). However, good diagnostic accuracy is needed for it to be useful in distinguishing DLD in bilingual children.

Guiberson and Rodriguez (2013) and Roy and Chiat (2004) used both percent phoneme correct and whole word scoring in their studies to compare the effectiveness of each. Roy and Chiat (2004) found that neither form of scoring was more informative than the other. The participants in their study were all typically developing, but they found that whole word scoring was sensitive to differences between the participants, such as age, and to the length of the nonword. Guiberson and Rodriguez (2013) found likelihood ratios that were moderately strong for whole word scoring and a little weaker for percent phoneme correct scoring. Also, for percent phoneme correct scoring they achieved adequate specificity but unacceptable sensitivity and the confidence interval ranges spanned into uninformative values. However, with whole word scoring, they found that there was adequate sensitivity and specificity and the confidence intervals did not include uninformative values. These studies show some preliminary data that whole word scoring may be as effective as or more effective than percent phoneme correct scoring. However, more research is necessary to back up these preliminary findings. Several studies that have used whole word scoring have used only typically developing children or only monolingual children. There are relatively few studies that use whole word scoring with bilingual children with and without DLD. This is a gap that needs to be filled with more research.

Error Patterns

Types of errors in NWR might offer an alternative to differentiation between the two groups. Few studies have looked at specific error patterns children make within the nonword repetition tasks. Ebert et al. (2008) looked for errors in their study done with all typically developing children. They separated the errors into three groups: consonant, vowel, and syllable errors. Syllable errors were errors on both the consonant and the vowel in the same syllable.

They found that almost 74% of errors made were syllable errors, and then approximately 21% being consonant errors and approximately 5% being vowel errors. However, even though there were relatively few vowel errors, more than half of the participants made an error on a vowel. This is in contrast to a study done by Girbau and Schwartz (2007) who found that typically developing children were almost perfect in their vowel production. They also found that children with DLD made more consonant cluster errors than typically developing children. Substitutions were the most frequent error type for both groups and the children with DLD made significantly more substitution and omission errors than the typically developing children. The group with DLD made more than three times as many substitution errors and six times as many omission errors as the typically developing group. Relatively few studies have looked at addition errors, partly due to the fact that when using Dollaghan and Campbell's (1998) scoring methods, additions are not counted as incorrect. Girbau and Schwartz (2007) compared addition errors made by children with language impairment and those without and found no group difference in the frequency of consonant additions.

Gathercole et al. (1994) looked at single phoneme and multiple phoneme errors. They found that single phoneme substitutions were the most common error made (26%). Other common errors were substitutions + deletions (22%), single phoneme deletions (16%), multiple phoneme substitutions (12%), and multiple phoneme deletions (12%). Roy and Chiat (2004) also looked at the types of errors children made, however they focused on prosodic implications, such as the stress of the syllables missed. They found that the majority of syllables deleted were unstressed syllables, that younger children deleted more syllables than older children, and that most syllables lost were in the three syllable items, which was the longest syllable length.

They found that overall doing a whole syllable loss analysis showed differences between children.

None of the studies mentioned above have had groups of children that are both bilingual and separated by language group (language impaired vs. typically developing). There has been preliminary evidence that error patterns can show differences between children, however more work needs to be done to determine if error patterns could be used to distinguish bilingual children with developmental language disorder from those with typical language.

Purpose of the Current Study

To address the identified gaps in the research, the current study examined English and Spanish NWR patterns in bilingual children with DLD. First, the effectiveness of whole word scoring and percent phoneme correct scoring in distinguishing between bilingual children with and without developmental language disorder was examined. Given previous findings (Guiberson & Rodriguez, 2013; Roy & Chiat, 2004) we hypothesized that whole word scoring would be as effective as or even more effective than percent phoneme correct scoring in determining which children have DLD. Second, this study compared the error patterns of bilingual children with developmental language disorder and those without to begin looking at the clinical utility of analyzing error patterns in a nonword repetition task. Given previous research (Ebert et al., 2008; Girbau & Schwartz, 2007; Gathercole et al., 1994), we expected to find that children with DLD would make significantly more errors than typically developing children in these areas: consonant, vowel, omission, and substitution errors. We expected to find that the largest discrepancy between the two groups would be in the number of vowel errors and omission errors. Previous findings (Girbau & Schwartz, 2007) did not find a difference between

groups for additions, however, there were relatively few studies that included additions in their error analysis.

Method

Participants

A sample of 26 Spanish-English bilingual children between the ages of 6;0-9;4 (mean: 8;0) participated in the study. They were recruited from three elementary schools and a university speech and language clinic in the southwest United States. Each parent filled out a consent form in their preferred language. All participants spoke both Spanish and English, well enough to tell a story in each language. Thirteen of the participants had developmental language disorder (DLD) and 13 had typically developing language skills. Participants with DLD were recruited first and then matched with typically developing participants by age (within 7 months), gender, and grade.

Inclusionary Criteria

To be included in the DLD group, participants needed to be receiving speech and language services at the time of the study. To be included in the TD group, participants could not be receiving speech and language services and they needed to score higher than one standard deviation below the mean on the Receptive One Word Picture Vocabulary Test: Spanish-Bilingual Edition (ROWPVT) (Brownell, 2001). All participants completed the ROWPVT. The mean for the DLD group was 95.62 and the mean for the TD group was 112.23. The difference between groups was significant ($p = <0.5$).

Exclusionary Criteria

Children were excluded from this study if their parents reported a history of hearing impairment, articulation problems, cognitive impairment, or social and emotional behavioral

problems. Each participant completed a hearing screening as part of the task battery for the study. Additionally, the participants all completed the Abbreviated Version of the Universal Non-verbal Intelligence Test (UNIT), which included the subtests for symbolic memory and cube design (Bracken & McCallum, 1998). Each participant had to score above 1.5 standard deviations below the mean to be included in the study. The mean score on the UNIT for the DLD group was 97.69 and the mean score for the TD group was 103.46. Overall, neither group scored significantly higher than the other on the UNIT (p = >0.5).

Socioeconomic Status

School lunch eligibility and maternal and paternal education levels were gathered for each participant. Twenty of the participants qualified for free lunch from school, four had regular lunches, and there was no data for two of the participants. Education levels are reported in three categories: up to high school, finished high school/GED, and beyond high school (see Table 1). For maternal education, the DLD group had six mothers who completed grades up to high school, two mothers who finished high school or received their GED, and five mothers who received education beyond high school. The TD group had six mothers who completed grades up to high school and six mothers who received education beyond high school. There was no information for one participant. For paternal education, the DLD group had four fathers who completed grades up to high school, three fathers who finished high school or received their GED, and two fathers who received education beyond high school. There was no information reported for four participants. The TD group had three fathers who completed grades up to high school, two fathers who completed high school or received their GED, and three fathers who received education beyond high school. There was no information for five participants.

Table 1

Participant Information

	Average Age	Gender (n)		Maternal Education (n)			Paternal Education (n)		
		Male	Female	Up to HS	HS/GED	Beyond HS	Up to HS	HS/GED	Beyond HS
DLD	8;1 (1;1)	9	4	6	2	5	4	3	2
TD	8;1 (0;11)	9	4	6	0	6	3	2	3

Note. DLD = Developmental Language Disorder. TD = Typically Developing. HS = High School. GED = General Educational Development.

Language Experience

The parents and teachers for each participant filled out the Bilingual Input Output Survey (BIOS) from the Bilingual English Spanish Assessment (BESA; Pena et al., 2014) to determine the language exposure and use for each child. Parents reported the child's input and output language experience on a daily and hourly basis. Input and output percentages were calculated following the procedures from the BESA and found in Table 2. For the DLD group, the mean for English input/output was 0.58 (or 58%) and the mean for Spanish was 0.42 (or 42%). For the TD group, the mean for English input and output was 0.48 (or 48%) and the mean for Spanish was 0.52 (or 52%). There was no information given for one student in the TD group.

Parents also completed the Inventory to Assess Language Knowledge (ITALK) from the BESA (Pena et al., 2014). They reported on how their child performs in different aspects of each language, including vocabulary proficiency, speech proficiency, sentence production proficiency,

grammatical proficiency, and comprehension proficiency on a scale of 1-5 with 1 being low and 5 being the highest, or most proficient. For the DLD group, the mean score for English was 3.85 and the mean score for Spanish was 3.65. For the TD group, the mean score for English was 4.24 and the mean score for Spanish was 3.95. There was no information reported for one student in the DLD group for English, two students in the TD group for English, and one student in the TD group for Spanish. There was no significant difference between the groups for Spanish or English (p = >0.5).

Table 2

Language Experience

	BIOS (Input/Output)		ITALK		Mean age of 1st exposure to English
	English	Spanish	English	Spanish	
DLD	0.58 (0.24)	0.42 (0.24)	3.85 (0.55)	3.65 (0.72)	3 (2)
TD	0.48 (0.18)	0.52 (0.18)	4.24 (0.59)	3.95 (0.64)	2 (2)

Note. BIOS = Bilingual Input Output Survey. ITALK = Inventory to Assess Language Knowledge. DLD = Developmental Language Disorder. TD = Typically Developing. Standard deviations are noted in parentheses.

Language Abilities of Groups

Participants also completed a sentence repetition task in English and Spanish from the Clinical Evaluation of Language Fundamentals-4 (Semel et al., 2003) and the Clinical Evaluation of Language Fundamentals-4, Spanish, (Wiig et al., 2006) respectively. Scores are found in Table 3. The sentence repetition subtest has a mean of 10 and a standard deviation of 3. The mean for the DLD group in English was 2.62 and the mean for the DLD group in Spanish was 3.77. The mean for the TD group in English was 5.92 and the mean for the TD group in Spanish

was 8.85. The TD group scored significantly higher than the DLD group in English and Spanish (p = <0.5).

Each participant also completed a narrative in English and Spanish using Mercer Mayer's *Frog Goes to Dinner* (1974) and *Frog, Where Are You?* (1969), which were randomly assigned. Percent grammatical utterances was calculated for the narratives by dividing the number of grammatical utterances by the total number of grammatical and ungrammatical utterances. The mean for the DLD group in English was 0.62 (SD=0.20) and the mean for the DLD group in Spanish was 0.62 (SD=0.22). The mean for the TD group in English was 0.77 (SD=0.13) and the mean for the TD group in Spanish was 0.80 (SD=0.13). The means for the TD group were significantly higher than the DLD group in both English and Spanish (p = <0.5).

Table 3

Test Scores

	UNIT	ROWPVT	SR		Grammaticality	
			English	Spanish	English	Spanish
DLD	97.69 (14.13)	95.62 (19.44)	2.62 (1.76)	3.77 (2.38)	0.62 (0.20)	0.62 (0.22)
TD	103.46 (13.13)	112.23 (16.74)	5.92 (2.40)	8.85 (3.16)	0.77 (0.13)	0.80 (0.13)

Note. UNIT = Universal Non-verbal Intelligence Test. ROWPVT = Receptive One Word Picture Vocabulary Test. SR = Sentence Repetition. DLD = Developmental Language Disorder. TD = Typically Developing. Standard deviations are noted in parentheses.

Nonword Repetition Measures

The participants all completed a nonword repetition task in English and Spanish. The English nonword repetition task came from Dollaghan and Campbell (1998). These nonwords were created to minimize word likeness and consists of 16 nonwords, with four nonwords at each

syllable length (one to four syllables). Each word begins and ends with a consonant and there are no consonant clusters or "late eight" consonant sounds, which helped control for articulatory errors. Only tense vowels are included, to make them easily distinguishable. The Spanish nonword repetition task was adapted from Gutiérrez-Clellan and Simon-Cereijido (2010). This list consisted of 16 nonwords, with four words at each syllable length (two to five). Words at the syllable lengths of two through four were taken directly from the Gutiérrez-Clellan and Simon-Cereijido (2010) list and then five syllable words were created by rearranging syllables from the list. These syllable lengths were included because they are the most common lengths in Spanish words. This list used only two later developing phonemes to control for articulation errors. Similar to the Dollaghan and Campbell (1998) list, syllables were used that had limited frequency in that position of the word.

Procedures

Each participant completed the tasks in a quiet room in their school or clinic. The tasks were all administered by a bilingual research assistant and instructions were given in the language being tested at the time. Each student completed all the tasks in one to three sessions ranging no more than 3 months in time from the first to the last session. The order of presentation of tasks varied amongst students, with most students completing the Spanish tasks before the English tasks.

In administering the nonword repetition task, the research assistant began by saying, "You are going to hear some silly words. Listen carefully to each word and then tell me what you heard. Let's practice." The students were then given two practice nonwords. If the child did not repeat these nonwords, they were given additional practice items. After the practice items, children began the test. They listened to a recording of each nonword through headphones. They

heard each word only one time and then were asked to repeat it. All their responses were audio recorded.

The NWR task was scored in two different ways: on a percent phoneme correct basis and a whole word basis. For the whole word scoring, the participant received a score of 0 or 1 depending on if they got everything right in the word or not. The whole-word scoring mostly followed the criteria used by Guiberson and Rodriguez (2013), except that additions were marked as incorrect in the current study. Distortions were not marked as incorrect. Percent phoneme correct scoring was based off of guidelines provided by Dollaghan and Campbell (1998). Substitutions and omissions were scored as incorrect, distortions were scored as correct, and additions were not counted as incorrect because they did not indicate a loss of information. When sounds were omitted, vowels acted as syllable anchors to line up the response to the rest of the target to maximize scores. The total number of correct phonemes was divided by the total number of phonemes for each language resulting in percent phonemes correct (PPC).

An error analysis was also done on the nonword responses. The types of errors made in the NWR task were counted by whether the errors were made on a vowel or a consonant (Ebert et al., 2008; Girbau & Schwartz, 2007). The number of additions, omissions, and substitutions (Gathercole et al., 1994; Girbau & Schwartz, 2007) made were also counted. In order to compare the rate of error types, the number of errors for each category was then divided by the total number of phonemes, except in the case of additions.

Types of errors were scored by using an adapted version of the Nonword Scoring Protocol (Gray et al., 2019). According to this protocol, the rater may "slide" phonemes or syllables over in the participant's answer to match the appropriate phonemes and syllables in the nonword to maximize the participant's points, as long as the rater does not change the order of

the phonemes. The rater may also "pop out" additional phonemes or syllables to maximize points and syllables or phonemes may hang off the beginning or end of the nonword. Syllable structure must be maintained, keeping the consonant and vowel of the syllable together (Gray et al., 2019). Vowels in the participant's response must be lined up with vowels in the nonword and consonants must be lined up with consonants. When there are several substitutions in the participant's response and it is difficult to match up to a specific syllable in the nonword, it should be lined up starting at the beginning of the nonword and moving forward. However, if a substituted phoneme is similar to one in the nonword (e.g., /b/ for /p/), the rater may slide the syllable over, as this is the closest substitution.

To determine interrater reliability for PPC scoring, a second bilingual research assistant listened to the audio recorded NWR tasks and scored 15.4% of the tasks. (3/26 for English NWR and 5/26 for Spanish NWR). The participants were randomly chosen to calculate interrater reliability. The interrater reliability ranged from 85.4% to 86.0%, with an average of 86.3% for the English NWR task. The interrater reliability ranged from 85.0%-90.0%, with an average of 86.8% for the Spanish NWR task. Whole word scoring and error analysis was completed after transcription of the nonwords and percent phoneme scoring. To determine interrater reliability for whole word scoring and error analysis, a second research assistant scored 19.2% (5/26) of the English and Spanish NWR tasks. The participants were randomly chosen to calculate interrater reliability. The average interrater reliability for the English NWR task was 97.3% (95.8-97.9%) and the average interrater reliability for the Spanish NWR task was 95.9% (94.7-97.0%).

Analysis

To address the first purpose of the study, Repeated Measures ANOVAs were performed with group (DLD, TD) as the between subjects independent variable and language (English, Spanish)

as the within subjects independent variable. NWR scores in percent phoneme correct and whole word scores served as the dependent variables. A Discriminant Function Analysis was performed to calculate the diagnostic accuracy of both scoring methods, including sensitivity and specificity. Likelihood ratios were calculated from sensitivity and specificity scores. Sensitivity and specificity measures were interpreted following Plante and Vance's (1994) guidelines. Ninety percent and above was considered to have good discriminant accuracy. Eighty percent and above was considered fair. Likelihood ratios were interpreted following Dollaghan's (2007) guidelines. A positive likelihood ratio of 10 or above was considered as *effective* to rule in a disorder. A positive likelihood ratio between 3 and 9.9 was considered to *suggest* a disorder but was not sufficient. A negative likelihood ratio of 0.10 or less was considered as *effective* in ruling out a disorder. A negative likelihood ratio between 0.11 and 0.30 was considered as *suggestive* to rule out a disorder but should be interpreted with caution.

To examine error patterns in NWR, Repeated Measures ANOVAs were performed with group and language as the independent variables and the following five dependent variables: percent of consonant errors, percent of vowel errors, number of addition errors, percent of substitution errors, and percent of omission errors.

Results

Scoring Method

Percent phoneme correct scores and whole word scores were analyzed to determine their significance. Mean scores are reported in Table 4. Both PPC and WW scores were statistically significant by group, $F(1, 24) = 16$, $p=0.001$, $\eta_P^2=0.400$ and $F(1, 24) = 7.593$, $p=0.011$, $\eta_P^2=0.240$, respectively. For both score types, children in the DLD group performed below the TD group. English PPC scores were significantly higher than Spanish PPC scores, $F(1, 24) =$

12.548, p=0.002, η_p^2=0.343. However, there was no statistically significant difference between English and Spanish WW scores, F(1.0, 24.0) = 2.006, p=0.170, η_p^2=0.077. There were no significant interaction effects for either method of scoring, F(1.0, 24.0) = .561, p=0.461, η_p^2=0.023 for PPC and F(1, 24) = 3.729, p=0.065, η_p^2=0.134 for WW.

Table 4

Means Scores

Percent Phoneme Correct, Whole Word, Percent Consonant Correct, and Percent Vowel Correct Means/Standard Deviations

	DLD	TD
English PPC	0.55 (0.09)	0.66 (0.11)
English WW	1.62 (1.32)	1.92 (1.55)
Spanish PPC	0.48 (0.07)	0.61 (0.08)
Spanish WW	0.62 (0.65)	2.08 (0.64)
English PCC	0.55 (0.10)	0.64 (0.09)
English PVC	0.52 (0.11)	0.66 (0.14)
Spanish PCC	0.42 (0.09)	0.57 (0.09)
Spanish PVC	0.46 (0.09)	0.59 (0.10)

Note. PPC = percent phoneme correct. WW = whole word. PCC = percent consonant correct. PVC = percent vowel correct. DLD = developmental language disorder. TD = typically developing. Scores reported as percentages for PPC, PCC, and PVC and number of words correct for WW. Standard deviations are noted in parentheses.

Sensitivity and specificity were also calculated for whole word and percent phoneme correct scoring in Spanish, English, and for Spanish and English scores together and can be found in Table 5. Spanish and Spanish/English whole word scoring had the highest sensitivity and specificity with 92.3% for sensitivity and 84.6% for specificity. Spanish PPC and Spanish/English PPC had a sensitivity of 84.6% and specificity of 76.9%. English PPC had a sensitivity of 69.2% and a specificity of 61.5%. English WW scoring had the lowest sensitivity

and specificity with 53.8% for both. According to the guidelines from Plante and Vance (1994), the only measure with good discriminant accuracy is the sensitivity rating for Spanish and Spanish/English whole word scoring. Specificity for Spanish and Spanish/English whole word and sensitivity for Spanish and Spanish/English PPC scores are considered fair according to these guidelines. Both whole word and percent phoneme correct scoring were significant for Spanish and Spanish/English together. PPC scoring was significant for English, however whole word scoring was not significant for English.

Likelihood ratios were also calculated from the sensitivity and specificity measures. As expected from the sensitivities and specificities, Spanish and Spanish/English whole word scoring had the best likelihood ratios, with a positive likelihood ratio of 5.99 and a negative likelihood ratio of .09. Spanish PPC and Spanish/English PPC was the next best measure with a positive likelihood ratio of 3.66 and a negative likelihood ratio of .20. English PPC had a positive likelihood ratio of 1.80 and negative likelihood ratio of .50 and English WW scoring had a positive likelihood ratio of 1.16 and a negative likelihood ratio of .86. According to Dollaghan (2007), Spanish and Spanish/English whole word scoring had a negative likelihood ratio that is considered effective in ruling out a disorder. Spanish and Spanish/English whole word scoring and Spanish and Spanish/English PPC both had a positive likelihood ratio that would be suggestive of a disorder and Spanish and Spanish/English PPC scoring had a negative likelihood ratio that would be considered suggestive to rule out a disorder.

Table 5

Sensitivity, Specificity, and Likelihood Ratios

	Spanish		English	
	PPC	WW	PPC	WW
Sensitivity	84.6%$^{+}$	92.3%$^{++}$	69.2%	53.8%
Specificity	76.9%	84.6%$^{+}$	61.5%	53.8%
+LR	3.66*	5.99*	1.80	1.16
-LR	.20*	.09**	.50	.86

Note. $^{+}$ fair and $^{++}$ good discriminant accuracy. * suggestive and ** effective in ruling in/out a disorder. +LR = positive likelihood ratio. –LR = negative likelihood ratio. PPC = percent phoneme correct. WW = whole word.

Error Analysis

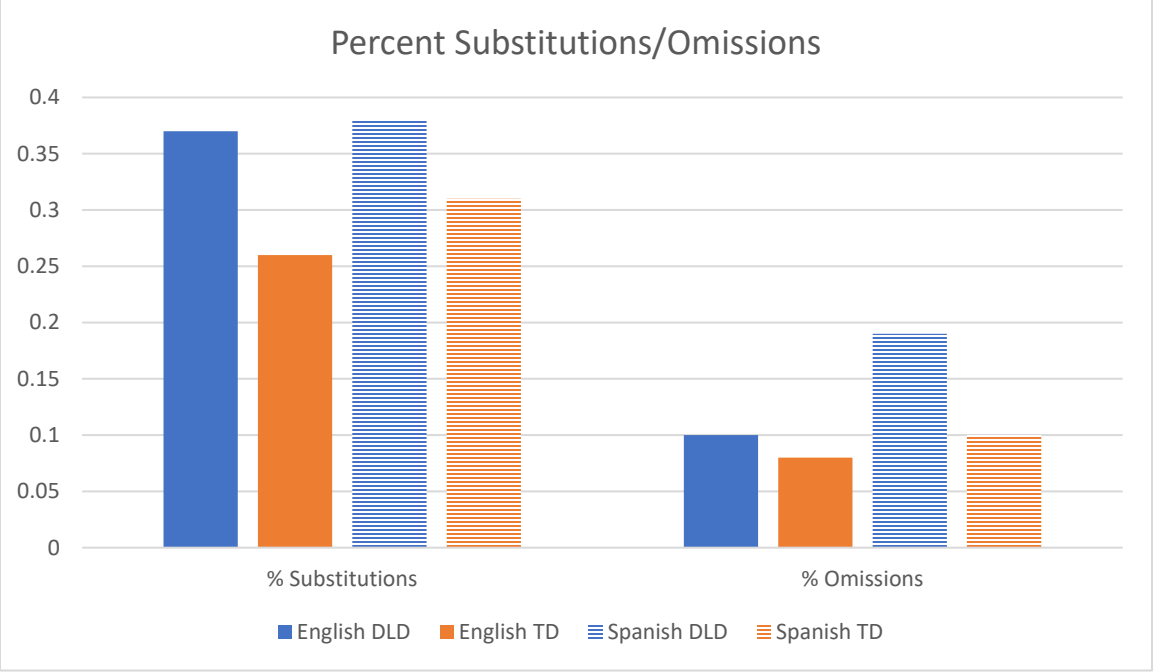
Both percent consonant correct and percent vowel correct were significant by group, $F(1, 24) = 17.192$, $p=0.000$, $\eta_p^2=0.417$ and $F(1, 24) = 11.392$, $p=0.003$, $\eta_p^2=0.322$ respectively. Children in the DLD group scored significantly lower than those in the TD group for both consonants and vowels correct (see Table 4). PCC and PVC were also significant by language $F(1.24)=27.498$, $p=0.000$, $\eta_p^2=0.534$ for PCC and $F(1, 24)=13.735$, $p=0.001$, $\eta_p^2=0.364$ for PVC. PCC and PVC scores were significantly higher in English than in Spanish. Neither PCC nor PVC had a significant group x language interaction $F(1, 24) = 2.118$, $p=0.159$, $\eta_p^2=0.081$, $F(1, 24) = 0.367$, $p=0.550$, $\eta_p^2=0.015$, respectively.

The percentage of errors for each type including additions, substitutions, and omissions were analyzed for significance. Additions were not statistically significant by group, language, or for group x language interaction effects ($p > .05$). The percentage of both substitution and

omission errors made were significant by group, $F(1, 24)=8.806$, $p=0.007$, $\eta_p^2=0.268$ for substitutions and $F(1, 24)=7.253$, $p=0.013$, $\eta_p^2=0.232$ for omissions. Children in the DLD group made significantly more substitution and omission errors than those in the TD group. Omissions were significant by language, $F(1, 24)=8.939$, $p=0.006$, $\eta_p^2=0.271$, with a significantly higher percentage of omission errors being made in Spanish than English. Substitutions were not statistically significant by language $F(1, 24)=2.668$, $p=0.115$, $\eta_p^2=0.100$. Neither substitutions nor omissions had statistically significant interaction effects between group and language, $F(1, 24)=1.165$, $p=0.291$, $\eta_p^2=0.046$ and $F(1, 24)=3.435$, $p=0.076$, $\eta_p^2=0.125$], respectively.

Figure 1

Percent Substitutions/Omissions



Note. DLD = developmental language disorder. TD = typically developing.

Discussion

The purposes of this study were to evaluate diagnostic accuracy of NWR tasks in English and Spanish and to explore error patterns. Three main findings emerged from the study. The first main finding was that both whole word (WW) scoring and percent phoneme correct (PCC) scoring were statistically significant in determining which group participants were in (TD or DLD), with the exception of whole word scoring in English. Spanish and Spanish/English whole word scoring had the best sensitivity, specificity, and likelihood ratios of any method of scoring. However, despite the seemingly good discriminant accuracy, whole word scoring may not be clinically useful. The mean score for Spanish WW scoring was 0.62 for the DLD group and 2.08 for the TD group, meaning that the average participant in the TD group was completely correct on only 2/16 nonwords and the difference between the two groups was small. Although statistically significant, these differences may not be clinically significant or practically useful. Previous studies also found whole word scoring to be statistically significant (Guiberson & Rodriguez, 2013; Roy & Chiat, 2004); however, there were some key differences in those studies. Roy and Chiat (2004) had all typically developing participants and only tested in English, using a list of real words and a nonword list that was different from the one used in this current study. Their participants were also younger than the participants used in the current study. They found that the repetition tasks were statistically significant in distinguishing participants by age. In looking at the mean scores, the older children had nearly five more words or nonwords correct than the younger children, while the standard deviation of the younger children was more variable. Guiberson and Rodriguez (2013) had participants in a TD and DLD group, however, they were predominately Spanish speaking and they only used a Spanish nonword repetition task. Their participants were also younger than the participants in the current

study (3;0-5;10). They found that whole word scoring had a higher diagnostic accuracy than PPC scoring. The participants in the Guiberson and Rodriguez (2013) study received much higher average WW scores than the participants in the current study, however, the difference between groups is still relatively small, with the TD group on averaging scoring less than four more nonwords than the DLD group. The Guiberson and Rodriguez (2013) study also used a different nonword list than the current study, indicating that perhaps the list used here may be more difficult than those used in other studies. This could explain the lower whole word correct scores for this study. Age could be another contributing factor. Perhaps younger children with DLD do not perform as well on tasks as older children with DLD and so the contrast between groups may be larger and have a greater impact on diagnostic accuracy.

As mentioned previously, NWR tasks are not meant to be the sole test in determining developmental language disorder in children (Guiberson & Rodriguez, 2013). Whole word scoring is useful because it can be done live and is more time efficient than PPC scoring (Guiberson & Rodriguez, 2013). Although the results in this study show it may not be practically useful to exclusively look at whole word scores, it could be used as a quick screener, able to rule out developmental language disorder if children score well above two whole words correct. Future research could be done using different NWR task lists to determine if whole word scoring would be more practical on those. Statistically speaking, it appears to be a good option.

The second main finding was that the Spanish NWR task had the best diagnostic accuracy. In both percent phoneme correct scoring and whole word scoring, the best sensitivity and specificity were achieved in Spanish scores or from the scores of Spanish and English together. Overall, the participants scored higher in English and the difference between DLD and TD children was less in English. Therefore, according to this study, English NWR scores did not

add discriminating information for the identification of language disorders. According to this study, the information gained from the Spanish NWR task alone would be sufficient to determine which children have DLD with the highest possible diagnostic accuracy. This could be due to the fact that the Spanish NWR task has longer nonwords than the English NWR task, as Spanish words are typically longer than English words (Ebert et al., 2008). However, with longer nonwords, the participants have to hold more in their working memory, which could be the reason for greater differentiation of DLD and TD. The longer, five syllable nonwords may be crucial to distinguishing between the two groups because it can look at more subtle differences (Guiberson & Rodriguez, 2013). The current study supports this claim that to distinguish differences, longer words may be necessary, to a length that is not achieved in English NWR tasks.

The third main finding was that in addition to the frequency of errors being different by group, omissions were also different by language. Consonants, vowels, substitutions, and omissions were looked at as a percentage in order to be able to compare errors across languages, as the Spanish NWR task had longer nonwords and more phonemes overall than the English NWR task. All of these error types (consonants, vowels, substitutions, and omissions) were statistically different by group, with the DLD children making more errors than the TD children. This outcome was expected. The percentage of omissions made were also statistically significant by language, with more omissions occurring in Spanish. This could be due to the longer nonwords in Spanish. There was more to remember, so more information was lost. The literature has suggested that children with DLD have difficulties with working memory (Gutiérrez-Clellan & Simon-Cereijido, 2010). These results support that claim, as more omissions were seen in the language with longer nonwords.

This study also looked at the difference in additions made by group. Dollaghan and Campbell (1998) set out in their scoring procedures that additions would not be counted as incorrect because it did not show a loss of information. Several other studies followed the same procedures for scoring set apart in Dollaghan and Campbell (1998), including Gutiérrez-Clellan and Simon-Cereijido (2010) and Ebert et al. (2008), so there was little information about additions being looked at for error patterns. The current study found no statistically significant difference by group or language for additions made. Therefore, it appears that, as the Dollaghan and Campbell (1998) scoring method suggests, looking at additions does not provide any additional information about language disorder.

Limitations and Implications for Future Research

The current study had some limitations. The sample size was relatively small with 26 participants, which could limit the generalizability of the findings. This study was also done as a retrospective analysis and was not designed to specifically compare scoring methods or find error patterns. The original scorers did not score with types of errors in mind, including counting additions as errors.

Future research studies could be conducted for scoring methods and error analysis with larger sample sizes for better generalizability. Studies could be conducted for scoring methods with different nonword lists to see if whole word scoring is of more practical use with different lists. Whole word scoring is a faster, more immediate process, so it would be beneficial to clinicians if it is found to have more practical use in other nonword lists. Future researchers could also determine cut-off scores with the highest specificity, sensitivity, and likelihood ratios in PPC and WW scoring for clinical use. Including a wider range of ages would also help disambiguate inconsistent findings in the current and previous studies.

Future research could also focus more on error analysis and examine whether error types and frequencies of errors can be used for diagnostic purposes and identification of DLD. Given the current study's findings of omission errors, future research could investigate the location of these errors within the nonwords and determine if the patterns differentiate between children with DLD and those without. Exploring the location of errors within the nonwords may also provide further insights into the language processing difference in children with DLD. Lastly, researchers might also examine relationships between types of errors made in a nonword repetition task errors in other tasks, such as grammatical or phonological errors to understand the difficulties experienced by bilingual children with DLD.

Clinical Implications and Conclusion

An important clinical implication of this study is that based on these findings, clinicians may only need to administer Spanish NWR tasks to find the highest diagnostic accuracy. This finding should be confirmed in future studies before a strong recommendation. But, having to test just one language in NWR would be a more efficient method of using NWR to diagnose DLD in bilingual children.

Two different scoring methods for nonword repetition tasks were examined to find the method with the best diagnostic accuracy. According to the results of this study, whole word scoring for the Spanish NWR task had the best diagnostic accuracy, however due to the small number of nonwords either group repeated correctly, it may not be practically useful. This study also found that the best diagnostic information was found with the Spanish NWR task. This could be practically useful if clinicians only need to administer and score one NWR task.

The errors made on both NWR tasks were also analyzed. Additions did not differentiate between children with DLD and those without. Children with DLD made significantly more

substitution and omission errors than those in the TD group. Significantly more omission errors were made in Spanish than English, which could be due to the longer nonwords in the Spanish NWR task. Perhaps these longer nonwords are necessary to fully distinguish between groups.

References

Adlof, S. M., & Patten, H. (2017). Nonword repetition and vocabulary knowledge as predictors of children's phonological and semantic word learning. *Journal of Speech, Language, and Hearing Research*, *60*(3), 682-693. https://doi.org/10.1044/2016_jslhr-l-15-0441

Baddeley, A. D. (1986). *Working memory*. Oxford University Press.

Bracken, B. A., McCallum, R. S. (1998). *Abbreviated Version of the Non-Verbal Intelligence Test (UNIT).* The Riverside Publishing Co.

Brownell, R. (2001). *Receptive One-Word Picture Vocabulary Test: Spanish-Bilingual Edition (EOWPVT-4 SBE).* Academic Therapy Publications.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment. *Journal of Speech, Language, and Hearing Research*, *40*(3), 519-525. https://doi.org/10.1044/jslhr.4003.519

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Paul H Brookes Publishing Company.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136-1146. https://doi.org/10.1044/jslhr.4105.1136

Ebert, K. D., Kalanek, J., Cordero, K. N., & Kohnert, K. (2008). Spanish Nonword repetition. *Communication Disorders Quarterly*, *29*(2), 67–74. https://doi.org/10.1177/1525740108314861

Federal Interagency Forum on Child and Family Statistics. (2020). *America's children in brief: Key national indicators of well-being, 2020*. Childstats.gov. https://www.childstats.gov/americaschildren/race.asp

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. https://doi.org/10.1017/s0142716406060383

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, *2*(2), 103-127. https://doi.org/10.1080/09658219408258940

Gibson, T. A., Summers, C., Peña, E. D., Bedore, L. M., Gillam, R. B., & Bohman, T. M. (2014). The role of phonological structure and experience in bilingual children's nonword repetition performance. *Bilingualism: Language and Cognition*, *18*(3), 551–560. https://doi.org/10.1017/s1366728914000248

Girbau, D. (2016). The Non-word Repetition Task as a clinical marker of Specific Language Impairment in Spanish-speaking children. *First Language*, *36*(1), 30–49. https://doi.org/10.1177/0142723715626069

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

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, *36*(2), 129–151. https://doi.org/10.1016/s0021-9924(03)00003-0

Gray, S., Fox, A. B., Green, S., Alt, M., Hogan, T. P., Petscher, Y., & Cowan, N. (2019). Working memory profiles of children with dyslexia, developmental language disorder, or both. *Journal of Speech, Language, and Hearing Research, 62*(6), 1839-1858.

Guiberson, M., & Rodríguez, B. L. (2013). Classification accuracy of Nonword repetition when used with preschool-age Spanish-speaking children. *Language, Speech, and Hearing Services in Schools*, *44*(2), 121-132. https://doi.org/10.1044/0161-1461(2012/12-0009)

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2010). Using Nonword repetition tasks for the identification of language impairment in Spanish-English-speaking children: Does the language of assessment matter? *Learning Disabilities Research & Practice*, *25*(1), 48-58. https://doi.org/10.1111/j.1540-5826.2009.00300.x

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research and Practice*, *21*(1), 19-29. https://doi.org/10.1111/j.1540-5826.2006.00204.x

Mayer, M. (1969). *Frog, where are you?* Dial.

Mayer, M. (1974). *Frog goes to dinner*. Dial.

Paul, R., Norbury, C., & Gosse, C. (2018). *Language disorders from infancy through adolescence: Assessment & intervention* (5th ed.). Elsevier Health Sciences.

Pena, E. D., Gutiérrez-Clellan, V.F., Iglesias, A., Goldstein, B. A., Bedore, L. M. (2014). *Bilingual English-Spanish Assessment (BESA).* Brookes Publishing.

Plante, E., & Vance, R. (1994). Selection of preschool language tests. *Language, Speech, and Hearing Services in Schools*, *25*(1), 15-24. https://doi.org/10.1044/0161-1461.2501.15

Roy, P., & Chiat, S. (2004). A prosodically controlled word and Nonword repetition task for 2- to 4-Year-Olds. *Journal of Speech, Language, and Hearing Research*, *47*(1), 223-234. https://doi.org/10.1044/1092-4388(2004/019)

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals, fourth edition (CELF-4).* Pearson. The Psychological Corporation/A Harcourt Assessment Company.

Summers, C., Bohman, T. M., Gillam, R. B., Peña, E. D., & Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language & Communication Disorders*, *45*(4), 480–493. https://doi.org/10.3109/13682820903198058

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*(1), 1–16. https://doi.org/10.1016/j.jcomdis.2012.08.002

Wiig, E. H., Semel, E., Secord, W. A. (2006). *Clinical evaluation of language fundamentals, fourth edition, Spanish (CELF-4 Spanish).* Pearson. The Psychological Corporation/A Harcourt Assessment Company.

Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-language Nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, *19*(4), 298-310. https://doi.org/10.1044/1058-0360(2010/09-0064)

APPENDIX A

Annotated Bibliography

Included in this bibliography are sources related to the usefulness of nonword repetition (NWR) tasks in diagnosing language impairment in preschool-school aged children. The studies in this bibliography also gave important evidence into the use of NWR tasks with bilingual children and the effect that different language exposures had on the results. It also includes studies using the Spanish and English NWR tasks that will be used in the current study and studies that have used multiple scoring methods for the NWR tasks and analyzed error patterns in the tasks, which are questions which will be addressed in the current study.

Adlof, S. M., & Patten, H. (2017). Nonword repetition and vocabulary knowledge as predictors of children's phonological and semantic word learning. *Journal of Speech, Language, and Hearing Research*, *60*(3), 682-693. https://doi.org/10.1044/2016_jslhr-l-15-0441

Objective: To determine the relative influence of nonword repetition and vocabulary knowledge on the phonological and semantic aspects of children's word-learning abilities. *Method:* 50 children (age 5-12) completed a task battery including a nonword repetition task (Comprehensive Test of Phonological Processing), a test of receptive vocabulary (Peabody Picture Vocabulary Test-4), expressive vocabulary (Expressive Vocabulary Test-2), and experimental word learning assessments (to assess phonological recall, phonological recognition, nonverbal semantic recall, verbal semantic recall, and semantic recognition). According to parent report, one child had ADHD, one had received services for learning English as a second language, and nine had received speech/language services.

The experimental word learning tasks were computerized and taught the children six pseudowords paired with an unfamiliar object referent along with features of each object. The children then completed five tasks to assess their knowledge of the words including naming (phonological recall), listening (phonological recognition), drawing (nonverbal semantic recall), describing (verbal semantic recall), and finding (semantic recognition) tasks.

Conclusions: The correlations between the predictor variables (age, NWR scores, and vocabulary scores) and the word learning assessments ranged from small to moderate. Age, nonword repetition, and vocabulary knowledge together accounted for up to 44% of the variance in word learning. The specific influences of nonword repetition and vocabulary varied across different word learning assessments. The nonword repetition task scores were stronger in predicting phonological recall, phonological recognition, and semantic recognition. Vocabulary knowledge scores were stronger in predicting verbal semantic recall. *Relevance to current study:* This study provides evidence that nonword repetition tasks can help predict children's word learning abilities. This would make it a valuable assessment of language abilities, which is an assumption of the current study.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136-1146. https://doi.org/10.1044/jslhr.4105.1136

Objective: To determine if the scores on a nonword repetition (NWR) task constructed to minimize word likeness could differentiate between children with typical language and those with language impairment. Also to determine the clinical utility of a nonword

repetition task in comparison to a norm-referenced language measure in distinguishing between children with language impairment and those without.

Method: Study 1: An English nonword repetition task was constructed to minimize word likeness. It included 16 total nonwords with four words at each syllable length (one to four syllables). Each word began and ended with a consonant and had no consonant clusters or "late eight" consonant sounds. Only tense vowels were included, which made it so there were no weak syllables in the nonwords. To reduce predictability, consonants were only in syllable positions where they occurred less than 25% of the time and no consonants or vowels were used more than once in a given nonword. 40 children (age 6;0 to 9;9), 20 with a language impairment and 20 with typical language, participated in a task battery consisting of a hearing screening, the Peabody Picture Vocabulary Test (PPVT), the Spoken Language Quotient of the Test of Language Development – Second Edition (TOLD-2), the Test of Non-Verbal Intelligence – Revised (TONI-R), a 10-minute conversational language sample, and the English nonword repetition task. The NWR task was scored by percent phoneme correct (PPC). Substitutions and omissions were scored as incorrect, distortions were scored as correct, and additions were not counted as incorrect because they did not indicate a loss of information. When sounds were omitted, vowels were used as syllable anchors to line up the response with the rest of the target to maximize scores. The scores of the two language groups were compared.

Study 2: 85 children, 40 from study 1 and an additional 45, (age 5;8-12;2) participated in this study. 44 had language impairment and 41 had typically developing language. The children were administered the same task battery as in study 1. The

nonword repetition task list was the same as in study 1 and scored in the same way (percent phoneme correct). The clinical utility of these measures was determined by calculating likelihood ratios. For the NWR task they looked at the three-syllable, four-syllable, and total percent phoneme correct scores.

Conclusions: Study 1: The typically developing group scored significantly higher on the three-syllable, four-syllable, and total percent phoneme correct scores than did the group with language impairment. For both groups, the four-syllable PPC was significantly lower than the other lengths and for the LI group the three syllable PPC was significantly lower than the two- and one- syllables. For the three-syllable, four-syllable, and total PPC there was no overlap between the two groups at the 99% confidence interval. Study 2: The likelihood ratio for ruling a child into the LI group based on a total percent phoneme correct score of 70% or lower was 25.15. This means that a score of 70% or lower on the total PPC is 25 times more likely to come from a child with language impairment than one without. This corresponds to a posttest probability of LI of more than 95%, so with a score of 70% or lower you could rule in the presence of a language disorder with a small likelihood of error. The likelihood ratio for ruling out a language disorder based on a total percent phoneme correct score of 81% or higher is .03. This means that a score of 81% or higher on the total PPC is $1/20^{th}$ as likely to come from a child with LI as one without. With a score of 81% or higher, you could rule out language impairment with a high degree of confidence. In contrast, the likelihood ratio for the Spoken Language Quotient of the TOLD-2 was not sufficient to identify a child with LI without further testing. It was also not able to rule out a child from having a

language impairment. Thus, results on the NWR task provided more accurate results in less time than those on the norm-referenced language measure.

Relevance to current study: This study is important as it shows that scores on a nonword repetition task can differentiate between children with language impairment and those with typical language. It shows that the list of nonwords created by this author can be effective at differentiating these groups of individuals. This same list of nonwords will be used in the present study.

Ebert, K. D. (2014). Role of auditory non-verbal working memory in sentence repetition for bilingual children with primary language impairment. *International Journal of Language & Communication Disorders*, *49*(5), 631-636. https://doi.org/10.1111/1460-6984.12090

Objective: To learn more about the role that nonverbal working memory plays in sentence repetition tasks. *Method:* 47 Spanish-English sequential bilingual children (age 5;6-11;2) with language impairment completed a task battery consisting of a nonword repetition task in English and Spanish, the Recalling Sentences subtest from the English and Spanish Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4), and a tonal pattern matching task to test nonverbal working memory. The English nonword list came from Dollaghan and Campbell (1998) and the Spanish nonword list came from Ebert et al. (2008). The NWR tasks were scored by percent phoneme correct (PPC) following the procedures of Dollaghan and Campbell (1998) and the scores reported in this study were the longest lengths from each language (four syllables for English and five syllables for Spanish).

Conclusions: After controlling for effects of age, they found that nonverbal working memory was significantly correlated with sentence repetition in English and

Spanish. The nonword repetition scores were significantly correlated with sentence repetition in English and Spanish. Nonverbal working memory was not significantly correlated with the nonword repetition tasks. Predicting sentence repetition performance using nonverbal working memory, nonword repetition, and age was significant in both languages. Nonverbal auditory working memory was significant after accounting for age and NWR performance in each language. This shows that sentence repetition performance cannot be fully explained without nonverbal skills. This supports other studies that have shown that children with a language impairment have weaknesses in nonverbal working memory skills and that this can lead to poor performance on sentence repetition tasks. *Relevance to current study:* This study gives more insight into factors influencing the performance of children with language impairment on various tasks. This task proves that NWR tasks can be used to predict other language skills.

Ebert, K. D., Kalanek, J., Cordero, K. N., & Kohnert, K. (2008). Spanish nonword repetition. *Communication Disorders Quarterly*, *29*(2), 67–74. https://doi.org/10.1177/1525740108314861

Objective: To create a Spanish nonword list and obtain preliminary data for a Spanish nonword repetition task using this list with typically developing children. *Method:* 14 Spanish speaking typically developing preschoolers (age 3;5-5;6) participated in a task battery consisting of a speech and hearing screening, the Preschool Language Scale – Fourth Edition (PLS-4) in Spanish and English, and a Spanish nonword repetition task. The children all spoke Spanish as a first language and had various exposure to English. The NWR list was created to follow the phonotactic properties of Spanish. Construction of syllables and the stress of the nonwords followed Spanish patterns. This made the

stimuli more wordlike. Syllables were constructed with a consonant-vowel pattern, which is common in Spanish. The frequency of sounds in Spanish was adhered to with the most frequent consonants appearing in more of the nonwords than the less frequent sounds. The nonwords ranged in length from one syllable to five syllables. There were 4 nonwords at each length, resulting in 20 total nonwords. The list adhered to the criteria by Dollaghan and Campbell (1998) as much as possible, with some exceptions. For example, 12 of the syllables ended up corresponding to true words and some nonwords included a phoneme more than once in the same word. Later developing consonants were not included and there were no consonant clusters. All vowels were tense. The NWR task was scored by percent phoneme correct according to the procedures by Dollaghan and Campbell (1998).

Conclusions: Accuracy decreased between the three- and four- syllable levels and the four- and five- syllable levels. The younger children (3;6-4;0) decreased in accuracy between the three- and four- syllable lengths, but the older group (4;3-5;6) maintained accuracy between those lengths and then declined at the five-syllable level. Errors were divided into three groups: consonant, vowel, and syllable errors. 73.7% of errors were made at the syllable level (an incorrect vowel and consonant), 21.4% of errors were consonant errors, and 4.9% of errors were vowel errors. However, even though vowel errors were relatively small, more than half of the participants made a vowel error. They found that the Spanish and English Expressive Communication subtest raw scores were significantly correlated, but the NWR scores did not significantly correlate to the English or Spanish PLS-4 scores. *Relevance to current study:* This study showed evidence that this nonword repetition task is appropriate to use with Spanish speaking children and is

sensitive to age. It also gives data about the error patterns made by typically developing children, which provides a background for the current study to look into the differences in errors made by typically developing and language impaired children.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. https://doi.org/10.1017/s0142716406060383

Objective/Method: This article gives a theoretical framework for nonword repetition tasks drawn from multiple studies looking at samples of typical children and adults and children with language disorders. *Conclusion:* The ability to repeat words is perhaps the most effective predictor of language learning skills. The ability to repeat nonwords depends on storage of phonological segments in short-term memory or the "phonological loop". Individuals vary on the quality of that storage. Articulatory suppression, phonological similarity, endurance of the phonological representations, and increased stimulus length can all impair phonological short-term storage. Studies have found that for people with deficits in phonological short-term storage, they could learn word-word pairs, but it was difficult to learn nonword-word pairs. If the quality of the temporary phonological representation in the phonological loops is compromised, it will result in slow learning rates.

Vocabulary scores and nonword repetition scores are linked, particularly during the early stages of acquiring a language when children cannot rely on previously learned lexical representations. Even in older children and adults, there is a link between NWR scores and the ability to learn new words. The link between nonword repetition and word learning is in the sound form of the word. Redintegration is a phenomenon where children can rely on lexical representations to reconstruct incomplete representations in

the phonological loop. Nonwords don't apply to redintegration because they are not lexically based. The more wordlike a nonword is, the easier it is to repeat it because language knowledge can play a role in the perceptual analysis and construction of phonological representations or it can overlap with lexical phonological representations. This is also true for NWR tasks based off of language rules similar to the one someone actually speaks. There is a time-based decay for things held in the phonological loop.

Children with language impairment struggle with NWR tasks and the acquiring new phonological forms. Children with LI will score lower on a NWR task than children who are matched based on language skills. Children may continue to show deficits in NWR tasks even if their language and vocabulary catch up to a typical level. Children with language impairment's repetition scores decline as syllable length increases, which may have something to do with the time-based decay.

Deficits in nonword repetition may have a genetic basis. NWR scores do not differ based on characteristics such as race or maternal education. NWR scores and memory span are linked. Auditory processing also influences NWR. There is a debate on whether phonological storage or phonological processing and sensitivity makes the difference for nonword repetition. There is also a link between nonword repetition and output processes. With there's a motor speech problem then it will either lead to inaccurate productions or it will increase the time it takes to repeat the nonword, which puts more load on storage. Generally, studies will control for these factors. Deficits in phonological storage may not be sufficient to completely explain poor scores on a nonword repetition task, as children with language impairment tend to have better skills with serial recall tasks. Beyond phonological storage deficits, children with language

impairment showed deficits specific to nonword repetition. Impairments in working memory or significant cognitive load are two possible factors that could explain this missing piece. *Relevance to current study:* This provides a key framework for understanding nonword repetition tasks and their use in identifying language impairment.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, *2*(2), 103-127. https://doi.org/10.1080/09658219408258940

Objective: To give an overview of findings from an evaluation of The Children's Test of Nonword Repetition (CNRep) and its use in assessing children's language skills. *Method:* This article provided data from previously conducted studies using The Children's test of Nonword Repetition (CNRep). This task has 40 nonwords, 10 for each syllable level (originally one to four, later revised to two to five syllables). It follows the phonotactic rules and syllable stress patterns of English. It also minimized articulatory demands. Whole word scoring is used with this measure with allowances made for dialectal differences or when a child consistently pronounces one phoneme as another.

In one study, the scores of children with language impairment on this test was compared to typically developing children of the same age and matched nonverbal intelligence test scores and younger children with matched language abilities. In a large study to determine normative data, the test was given to 612 children between 4-9 years old in England. To determine test-retest reliability 63 5-year old children and 25 7-year old children completed the test twice. To study error patterns in the test, they studied the responses of 27 4-year old children who scored at the high end or low end of the test. They divided the errors into single phoneme categories and multiple phoneme categories.

To compare CNRep scores to other language abilities, one study looked at 4-5 year old children and another study looked at 8-year old children who completed the CNRep, an auditory digit span task, the Short Form of the British Picture Vocabulary Scale (BPVS), and Reading Test A of the British Abilities Scales (BAS). The 4-5 year old children also completed the Test for the Reception of Grammar (TROG).

Conclusions: In one study, the scores on the NWR task perfectly discriminated between the children with language impairment and those without. In the normative study, they found that scores on the test increased with age, except between 8 and 9 years old. However, other studies have found the test to be sensitive to differences in language abilities up to 10 years old. In each age group (by year), the scores went down for the four-syllable words and then improved slightly from there for the five-syllable words. This change could be due to the fact that several five-syllable words had grammatical morphemes familiar to children (such as -atory, -ually, etc.). The test-retest reliability was satisfactory.

In analyzing error patterns they found that single phoneme substitutions were the most common error (26%). Other common errors were substitutions + deletions (22%), single phoneme deletions (16%), multiple phoneme substitutions (12%), and multiple phoneme deletions (12%). Scores on the CNRep and the auditory digit span task were significantly correlated. There was a significant correlation between performance on the NWR task and receptive vocabulary scores, especially for the 4-5 year olds. CNRep scores were more effective in predicting vocabulary scores than the digit span task in 4-5 year olds. Between the ages of 4-5 the children's skills on the NWR task had a strong causal influence on vocabulary knowledge. CNRep scores were also significantly

correlated with the BAS reading scores for the 5 and 8 year old children, and the TROG scores measuring comprehension for 4 and 5 year olds. Nonword repetition tasks rely on phonological working memory, long-term knowledge, phonological analysis, and output processes. *Relevance to current study:* This article gives several examples of the relevance of nonword repetition tasks in identifying language impairment. It also focuses on error patterns, which is a question this study will address.

Gibson, T. A., Summers, C., Peña, E. D., Bedore, L. M., Gillam, R. B., & Bohman, T. M. (2014). The role of phonological structure and experience in bilingual children's nonword repetition performance. *Bilingualism: Language and Cognition*, *18*(3), 551–560. https://doi.org/10.1017/s1366728914000248

Objective: This study looked at how language experience and the phonological structure of a language influenced bilingual children's performance on a nonword repetition task. *Method:* 52 typically developing kindergarteners participated in this study. Parent and teacher questionnaires were given to determine hour by hour language use of each child. Children had at least 20% input from each language according to these surveys. The participants were matched by inverse exposure to Spanish and English within 10% and within 1 year of first exposure to English. Thus, there were 26 pairs in two groups, an English dominant group and a Spanish dominant group. These children completed an English and Spanish nonword repetition task and the Bilingual English Spanish Assessment (BESA). The English nonword repetition task came from Dollaghan and Campbell (1998) and the Spanish nonword repetition task came from Calderon (Gutiérrez-Clellan). Only the two to four syllable lengths were included in analysis. The

NWR task was scored by percent phoneme correct. On the BESA, only the semantics tasks were reported in this study.

Conclusions: The Spanish dominant group scored higher on both the English and the Spanish NWR tasks than the English dominant group did. This could be due to a practice effect, as Spanish has longer words, in general, than English, so these children are used to remembering longer words. Children also scored higher on the Spanish NWR task than the English NWR task at all syllables, however it was only statistically significant at the four-syllable level. In the English NWR task, the scores on the two-syllable length were significantly different than for the four-syllable length, but there were no significant differences between the three-syllable length and the other lengths. In Spanish, there was a statistically significant difference in scores between the two- and three- syllable lengths, but then it plateaued and there were not significant differences between the three- and four- syllable lengths or the two- and four- syllable lengths. The effect sizes in these analyses ranged from medium to large. This shows that small differences between the languages' phonological structures and the children's different experiences with the languages had a large effect on their productions. Overall, language experience, current language exposure, and phonological structure had effects on NWR scores for bilingual children. *Relevance to current study:* This study shows that children perform differently based on their language experiences. We will expect to see more patterns like this in the current study as we look at another group of Spanish-English bilingual children.

Girbau, D. (2016). The Non-word Repetition Task as a clinical marker of Specific Language Impairment in Spanish-speaking children. *First Language*, *36*(1), 30–49. https://doi.org/10.1177/0142723715626069

Objective: To determine if the scores on a Spanish nonword repetition task can serve as a diagnostic marker to identify Spanish-speaking children with and without language impairment. To explore the relationship between performance on a nonword repetition task and other language measures to begin to develop better guidelines for assessment and intervention for this population. *Method:* 20 children with language impairment (ages 8;0-9;11) and 20 age matched children with typical language (ages 8;1-10;3) from Spain completed a task battery consisting of a hearing screening, a parent interview/SES scale, the Test of Nonverbal Intelligence (TONI-2), eight language tests/subtests (Peabody Picture Vocabulary Test (PPVT-III), Token Test for Children (TTFC-2), Test de Comprension de Estructuras Gramaticales (CEG), Wechsler Intelligence Scale for Children (WISC-IV) Vocabulary subtest, four Illinois Test of Psycholinguistic Abilities (ITPA) subtests: auditory comprehension, auditory association, verbal expression, grammatical integration), and a Spanish nonword repetition (NWR) task. The NWR task was the list developed by Girbau and Schwartz (2007). It was scored by whole word scoring and they analyzed it for the total percent correct and the percent correct from just the three- to five- syllable lengths.

Conclusions: There was a length effect, with most of the errors on the NWR task happening in the three- to five- syllable range. The language impairment (LI) group scored significantly lower than the typically developing (TD) group on the total PPC and the difference was twice as large in the three- to five- syllable composite score. This test

had good sensitivity and specificity and good likelihood ratios. For children with typical language, if they score higher than 50% on the three- to five- syllable composite score, the likelihood ratio is 0.00 and the negative predictive value is 100%. For children with language impairment, if they score equal to or lower than 50% on the three- to five-syllable composite score, the likelihood ratio is 6.67 and the positive predictive value is 86.96%. This means that if a child scores above 50% it is significantly likely that they do not have language impairment and if they score below 50%, it is significantly likely that they do. No child with language impairment in this study scored more than 50%. Sensitivity was 100% and specificity was 85%. This would indicate that this task is a good diagnostic measure for initial identification of language impairment. The scores on the language measures correlated significantly with the nonword repetition scores. They were significant for the full sample of children, but not for the groups separately. Overall, the pattern was the higher the child scored on the nonword repetition task, the higher they scored on the language test. The associations were moderately strong to strong. This provides some evidence that improving phonological working memory could improve these language skills and vice versa. *Relevance to current study:* This study shows that Spanish nonword repetition task scores can differentiate between Spanish-speaking kids with language impairment and those with typical language with a high level of accuracy. This study also found that whole word scoring is an effective way to score an NWR task. This will be the scoring used in the current study.

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

Objective: To create a list of nonwords that follow the phonotactic patterns of Spanish with Spanish prosodic patterns and syllables. Also to examine the relationship between performance on the nonword repetition task and other language measures. *Method:* 22 children from Spain (age 8;3-10;11), 11 with language impairment and 11 age and gender matched peers with typical language participated in this study. All children completed the Clinical Evaluation of Language Fundamentals-4 (CELF-4), the Spanish adaptation of the Illinois Test of Psycholinguistic Abilities (ITPA), and a hearing screening. The children in the LI group completed the Wechsler Intelligence Scale for Children-Revised (WISC-R) and five of them additionally completed the Test of Nonverbal Intelligence (TONI-2). The typically developing kids completed the Batería de Aptitudes Diferenciales y Generales (BADYG E2). Parents also completed a questionnaire indicating the extent to which Spanish was the child's primary language, their socioeconomic status, family history of language deficits, and any history of neurological disorders or behaviors characteristic of autism.

A nonword repetition task was constructed following Spanish phonotactic patterns in the syllable structure and segments included. The list is comprised of 20 nonwords, four words at each syllable length (one to five syllables). All nonwords began with a consonant and 12 included at least one consonant cluster. No high frequency syllables were used. Each syllable contained only one vowel and no diphthongs were used. Most sounds in Spanish were included in the task and the stress varied across different syllable positions. The NWR task was analyzed segment by segment for vowels, consonants (including the two in each cluster), and clusters. Each segment was categorized as correct, or as a substitution, omission, or addition.

Conclusions: Most errors occurred at the three-, four-, and five- syllable lengths. The language groups differed for the percentage of total nonwords correct, the three- to five- syllable composite, and the three-, four-, and five- syllable lengths separately. The number of nonwords produced correctly decreased as the syllable length increased. Vowel errors were rare, especially for the TD children who had near perfect performance. The children with language impairment (LI) made more consonant and cluster errors than the typically developing (TD) children. Substitutions were the most frequent error type for both groups. The LI group made significantly more substitutions and omissions than the TD group, with the LI group's substitution errors being more than three times higher than the TD group and omissions being six times higher. There was no group difference for additions. For children in the TD group, if they score above 50% for the three- to five-syllable composite the likelihood ratio is 0.00 with a negative predictive value of 100%. For the LI group, if they score lower than or equal to 50% on the three- to five- syllable composite, the likelihood ratio is 11.00 with a positive predictive value of 91.67%. This cutoff score of 50% accurately discriminates between children with LI and those without. The posttest probability is 91.67%. The sensitivity is 1.00 and the specificity is 0.91. The mean of the four ITPA subtests highly correlated with the total and the three- to five-syllable composite NWR scores. Individually, the Auditory Association and Grammatical Integration subtests correlated with the NWR scores. No significant association was found between NWR scores and the tests of intelligence. *Relevance to current study:* This study shows evidence that a Spanish NWR task can discriminate between children with LI and those without. This study also looked into some of the errors made by children

with LI and their typically developing peers, which will be expanded on in the current study.

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, *36*(2), 129–151. https://doi.org/10.1016/s0021-9924(03)00003-0

Objective: To determine if performance on a nonword repetition task and digit span task could distinguish between preschool aged children with language impairment and those without and to compare their diagnostic accuracy and a norm-referenced language test. To test the reliability and validity of a nonword repetition task. To determine if word repetition practice had an effect on nonword repetition scores. To determine if there was a significant difference in performance on 2 forms of an NWR task. *Method:* 44 children (4;0-5;11), 22 with language impairment (LI) and 22 age and gender matched peers with typically developing (TD) language completed a digit span task and two nonword repetition tasks. The first day they completed the digit span task and 1 nonword repetition task. The second day they completed the digit span task and the same nonword repetition task, but with a different word order. In days three through six the children played games in pairs and practiced repeating English words from one to five syllables in length. One week after the second NWR administration, the children completed the digit span task again along with the second nonword repetition list.

The nonword repetition task came from the Children's Test of Nonword Repetition developed by Gathercole et al. (1994). There were 40 nonwords across four syllable lengths (two to five). The words were randomly divided into two lists of 20

nonwords to use as the alternate forms of the NWR task. The nonword repetition task and digit span task were scored by whole word scoring. The children also completed the Structured Photographic Expressive Language Test-2 (SPELT-II), the Bankson-Bernthal Test of Phonology (BBTOP), and the language impaired group completed the Peabody Picture Vocabulary Test-Third Edition (PPVT-III) several days prior to the digit span and NWR tasks.

Conclusions: In all three nonword repetition tasks and digit span tasks the typically developing group scored significantly higher than the language impairment group. Performance on the NWR task and the digit span task was significantly correlated. For the NWR task both groups improved from administration 1 to 2 and then declined slightly from administration 2 to 3. On the digit span task, both groups improved from administration 1 to 2 then the LI group declined slightly from administration 2 to 3, but the TD group improved. More children in the LI group improved their scores from administration 1-2 than children in the TD group in both tasks, which shows there may be a differential practice effect for the children with LI. Sensitivity and specificity were high on the NWR task for the first administration (95% sensitivity, 100% specificity), lowered a little (but were still high) for the second administration, and specificity dropped more in the third administration. Digit span was less accurate at discriminating between the groups than NWR was, but still had high sensitivity on the first and thirdadministrations.

Both the NWR task and the digit span task had higher sensitivity overall than the SPELT-II. For test-retest reliability, NWR scores were significantly correlated for the LI group with each administration but were not significantly correlated for the TD group, except between time 1 and 3. Digit span scores were significantly correlated with each

administration for both groups. The speech and language test scores were not significantly correlated with the phonological test scores for the LI group, except for the SPELT-II and PPVT-III with the NWR task on the third administration. The relationship between the speech and language tests and the phonological measures was stronger for the TD group. *Relevance to current study:* This study provided more evidence that NWR tasks may be better than norm referenced language measures in identifying children with LI. They used whole word scoring and found that it had accurate discriminate accuracy for the NWR task. The current study will also be using whole word scoring and looking at its diagnostic accuracy.

Guiberson, M., & Rodríguez, B. L. (2013). Classification accuracy of Nonword repetition when used with preschool-age Spanish-speaking children. *Language, Speech, and Hearing Services in Schools*, *44*(2), 121-132. https://doi.org/10.1044/0161-1461(2012/12-0009) *Objective:* To describe and compare Spanish nonword repetition task performance between Spanish-speaking preschool age children with language impairment and those with typical language across two scoring methods. To contrast the classification accuracy of the two scoring approaches. *Method:* 44 predominantly (≥80% of the time) Spanish-speaking preschool aged children (3;0-5;10), 21 with language impairment and 23 without, completed a Spanish nonword repetition task (NWR) and the Spanish Preschool Language Scale-4 (SPLS-4). The NWR list was from Ebert et al. (2008). All testing was done in Spanish by a bilingual speech-language pathologist. Parents completed a questionnaire that included questions about family language usage patterns. The NWR task was scored by percent phoneme correct based on the Dollaghan and Campbell (1998) approach and by item-level scoring.

Conclusions: Age was significantly correlated with NWR scores, with older children scoring better than younger children. There was a significant effect for language impairment status and syllable length. As the syllables got longer, the scores went down. The language impaired kids especially struggled with three, four, and five syllables across both scoring methods, so the scores from these syllables were used in analyses. The language impairment (LI) group performed significantly lower than the typically developing (TD) group on the item level scoring. This shows promising results that item level scoring can show group differences in language skills. Item-level scoring had good sensitivity (71%) and specificity (74%) with moderately strong likelihood ratios (positive likelihood ratio was 2.74 and the negative likelihood ratio was .39). The confidence interval ranges did not include uninformative values. This indicates fair discriminant accuracy. For item level scoring the pretest probability was 48% and the positive posttest probability was 71%. This means that there is a 71% probability that children with LI will score poorly on three, four, and five syllables. This indicates that item-level scoring may help in indicating LI in Spanish speaking preschoolers.

Percent phoneme correct scoring had adequate specificity (78%), but unacceptable sensitivity (48%) with likelihood ratios weaker than that for item-level scoring (the positive likelihood ratio was 2.19 and the negative likelihood ratio was .67.) The confidence interval ranges went into uninformative values. The five-syllable nonwords were important to distinguish between groups. *Relevance to current study:* This shows some evidence that item-level scoring may have better classification accuracy than PPC scoring, which will be reviewed further in the current study.

Gutiérrez-Clellen, V. F., & Simon-Cereijido, G. (2010). Using Nonword repetition tasks for the identification of language impairment in Spanish-English-Speaking children: Does the language of assessment matter? *Learning Disabilities Research & Practice*, *25*(1), 48-58. https://doi.org/10.1111/j.1540-5826.2009.00300.x

Objective: To evaluate the clinical utility of a nonword repetition task for distinguishing between Spanish-English bilingual kids with and without language impairment. To determine how differences in language skills and use between their two languages affect the differentiation of these children. *Method:* 144 Spanish-English bilingual children (95 with typical language, 49 with language impairment) between the ages of 3;11-7;10 participated in this study. Parents and teachers filled out a questionnaire to determine language dominance. There were 63 Spanish dominant children (22 with LI and 41 without), 71 English dominant children (24 with LI and 47 without), and 10 where neither language was dominant. The children completed an English and a Spanish nonword repetition task.

The English nonwords came from Dollaghan and Campbell (1998) and the Spanish nonwords were created by these authors. The Spanish nonword list had 20 nonwords across three syllable lengths (two, three, and four). It followed the prosodic and phonological characteristics of Spanish. It was not based on the same sounds as the English NWR task, but only used Spanish consonants and vowels. The syllable lengths were chosen because they are the most common Spanish word lengths (two, three, and four syllables). Any nonwords that resembled English words were discarded. Similar to the Dollaghan and Campbell (1998) English nonwords list, syllables were used that had limited frequency in that position of the word. Only two late developing phonemes were

used, however the children were not penalized on these if they could not produce them correctly in spontaneous language. The NWR tasks were scored by percent phoneme correct, following the Dollaghan and Campbell (1998) procedures. Articulation errors were not penalized and Spanish influenced errors made on the English nonword repetition task were not penalized.

Conclusions: The kids with typical language scored significantly higher on both the English and Spanish nonword repetition tasks. The English NWR task had fair test accuracy. It had moderate specificity (0.82) but poor sensitivity (0.55). 45% of children with LI were not identified as such because they scored above the 70% cutoff score. The Spanish NWR task had moderate specificity (0.82) and inadequate sensitivity (0.61). Neither NWR task alone could accurately classify children with or without a language impairment. Several children with typical language did not pass the NWR task in both languages, probably because of different levels of experience with each language. When the two tests were combined, the specificity increased to 0.95, which is a good classification rate. For the cutoff score of 70%, there was an intermediate high positive likelihood ratio of 9.71. This means scoring 70% or lower on both the Spanish NWR task and the English NWR was more than nine times as likely to be a child with language impairment. Some English dominant children both with and without language impairment passed the Spanish NWR task and some Spanish dominant children, both with and without language impairment, passed the English NWR task. This shows that testing only one language, even the child's dominant language may not classify children accurately, so bilingual children should be tested in both languages. *Relevance to current study:* This study further shows the clinical utility of nonword repetition tasks. It

specifically supports the idea of using both English and Spanish nonword repetition tasks for Spanish-English bilingual kids. This study also used the same Spanish nonwords that will be used in the current study.

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research and Practice*, *21*(1), 19-29. https://doi.org/10.1111/j.1540-5826.2006.00204.x

Objective: To determine if two language processing measures (the Competing Language Processing Task and a Nonword Repetition Task) can distinguish between children with language impairment and those without regardless of their language experience, whether monolingual or bilingual. *Method:* 100 children (age 7;10-13;11) participated in this study. They were split into three groups: 28 monolingual English children with language impairment, 50 monolingual English children with typical language, and 22 Spanish-English bilingual children with typical language. The bilingual kids all learned Spanish as their first language and had between 4-8 years of experience learning English. The children all completed the Competing Language Processing Task (CLPT) and an English nonword repetition task (NWR). The CLPT task requires participants to listen to sentences in pairs, state yes or no whether they are true or not, and then recall the last word in each sentence. The Dollaghan and Campbell (1998) NWR list was used. The CLPT was scored as percentage correct for comprehension and recall and the NWR task was scored by percent phoneme correct following the procedures from Dollaghan and Campbell.

Conclusions: Age accounted for 17% of the total variance in the children's mean for CLPT Recall, but it did not account for any significant variation in the NWR

performance. The LI's group comprehension score on the CLPT was comparable to the typically developing children scores. The language impairment (LI) group had significantly lower performance than the monolingual typically developing (TD) group on the recall portion. The monolingual TD and the bilingual TD groups did not have statistically significant differences in scores for the recall portion and the monolingual LI and bilingual TD groups came close, but also did not reach a statistically significant difference in performance. On the nonword repetition task, the monolingual TD group scored significantly better than the LI group and the bilingual group. Effect sizes for the differences between groups were large. The bilingual group scored significantly better than the LI group. There was significant overlap in scores for the three groups on the one-, two-, and three- syllable nonwords, but the scores on the four-syllable nonwords separated the groups.

The likelihood ratios for the CLPT Recall scores were indeterminate when considering on the monolingual children and both the bilingual and monolingual children. For the NWR task, the cutoff score was 76%. For the monolingual groups, a child scoring ≤76% was 10.7 times more likely to be from the LI group. The posttest probability was 85%, which has intermediate high diagnostic power. A score of ≥93% could rule out language impairment with a high degree of confidence for a monolingual child with a likelihood ratio of 0.08 and a posttest probability of 4%. Adding the bilingual group into the calculations lowered sensitivity, but not specificity. A score of ≤72% was 5.07 times as likely to come from a child with language impairment. The posttest probability was 66%. The score and likelihood ratio for ruling out language impairment was the same. So this NWR task could rule out language impairment for all the children involved, but

could not rule it in. *Relevance to current study:* This provides evidence that a nonword repetition task that measures processing is better able to identify children with a language impairment than a storage task, such as the CLPT. It also provides evidence that a NWR task in one language alone may not be sufficient for use with bilingual children.

Roy, P., & Chiat, S. (2004). A prosodically controlled word and Nonword repetition task for 2- to 4-Year-Olds. *Journal of Speech, Language, and Hearing Research*, *47*(1), 223-234. https://doi.org/10.1044/1092-4388(2004/019)

Objective: To develop a word and matched nonword repetition task for younger children (age 2-4) that can identify differences between them. To compare performance on a receptive vocabulary test to the performance on this word/nonword repetition task.

Method: 66 typically developing children (2;0-3;11) from England completed a standardized receptive vocabulary test (the long form of the British Picture Vocabulary Scale) and a repetition task consisting of 18 words and 18 nonwords. The repetition task had six words and six nonwords at each syllable length (one to three). The stress patterns were manipulated throughout the measure and the words and nonwords were phonologically matched. The repetition task was scored by whole-word scoring, allowing for some articulatory difficulties, sociolinguistic variation, and substitution of phonetic variants. For comparison the task was also scored by percent phoneme correct scoring with a liberal phoneme score (articulatory and sociolinguistic variations allowed) and a conservative phoneme score (variations not allowed). They also calculated the syllables lost in the repetitions, either when a vowel was omitted (with or without neighboring consonants) or when two syllables were coalesced (combining the consonant of 1 syllable with the vowel of another).

Conclusions: Neither type of scoring appeared to be more informative than the other, so whole word scoring was used in the statistical analyses. The older children (3;0-3;11) scored significantly higher on the repetition task than the younger children (2;0-2;11). The younger children's scores were more variable than the older children. Overall, the group correctly repeated significantly more words than nonwords, showing a word status effect. Scores decreased as words and nonwords got longer. For the real words, the difference in scores from the two-syllable length to the three-syllable length was not significant, however it was significant for the nonwords. Whole syllable loss was relatively rare, but the majority of syllables lost were unstressed syllables, especially the unstressed syllables that came prestress. Unstressed syllables that came prestress were three times more likely to be omitted than those that came poststress and 40 times more likely to be omitted than stressed syllables. The younger children omitted just over twice as many syllables as the older children and syllables were most frequently lost in the three-syllable items. However, this is also relative to the type of syllable, with prestress syllables in two-syllable items being more vulnerable than poststress syllables in three-syllable items, but prestress syllables in three-syllable items being the more vulnerable than prestress syllables in two-syllable items.

Repetition scores overall and separated for words/nonwords were significantly correlated with receptive vocabulary scores (using the raw scores from the BPVS) and age. The NWR task accounted for a significant amount of change in vocabulary scores. There is evidence that whole word scoring and whole syllable loss analyses do show differences between children and could be used instead of a percent phoneme correct score. *Relevance to current study:* This study addressed aspects of both questions in the

current study. They used whole word scoring and found that it was as informative as percent phoneme correct scoring. They also looked at error patterns the children made, specifically with prosody. These questions will be looked at further in the current study with a sample of children with and without language impairment for comparison.

Summers, C., Bohman, T. M., Gillam, R. B., Peña, E. D., & Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language & Communication Disorders*, *45*(4), 480–493. https://doi.org/10.3109/13682820903198058

Objective: To examine the performance of children with varying exposures to English and Spanish on a nonword repetition task. To compare scores on an NWR task to measures of semantics and morphosyntax. *Method:* 62 children (4;6-6;5) participated in this study, with data reported for 60 of them. 54 of the children were tested in Spanish and English, one in Spanish alone, and five in English alone. The children completed a Spanish and English nonword repetition task, semantic task, and morphosyntax task. Parents filled out questionnaires to determine each child's current exposure to Spanish and English and the first year of English exposure. The semantic and morphosyntax task came from the Bilingual English Spanish Assessment (BESA). The English NWR task came from Dollaghan and Campbell (1998) and the Spanish NWR task came from Calderon (2003). The NWR task was scored by percent phoneme correct (according to the procedures by Dollaghan and Campbell) within each syllable length to control for the different numbers of nonwords in each task.

Conclusions: NWR scores declined as syllable length increased. On the English NWR task, the difference in mean between the two- and three- syllable lengths and the

four-syllable length was significantly different. On the Spanish NWR task the differences in mean between the two-, three-, and four- syllable lengths were all significantly different. Children's performance on the Spanish and English nonword repetition tasks was significantly correlated with their cumulative language experience and with their scores on the morphosyntax tests in both languages, but not with their scores on the semantic tests. Children's performance on the Spanish and English NWR tasks was similar, but the performance accuracy was higher in Spanish overall. Later exposure to English was correlated to a higher percentage on the four-syllable English nonwords and the three- and four- syllable Spanish nonwords, but a lower percentage on two- and three-syllable English nonwords. These findings may be due to the multisyllabic nature of Spanish and children's experience in saying longer words. Cumulative language experience seems to be more important for NWR performance than current language experience. Children had more experience and exposure overall with Spanish and they scored higher overall on the Spanish morphosyntax and semantics tests. Differences in NWR scores were not dependent on the test, but on the child's language experience. *Relevance to current study:* This study focused on how differing language exposure will affect the scores on an NWR test, which is an important factor when working with bilingual children, which this current study will do.

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*(1), 1–16. https://doi.org/10.1016/j.jcomdis.2012.08.002 *Objective:* To determine the effect of various levels of bilingual language exposure on a nonword repetition task and a sentence imitation task in French-English bilingual

children. To determine the diagnostic accuracy of the French version of these measures compared to a receptive vocabulary test with monolingual and bilingual children with and without language impairment. *Method:* Study 1: 84 5-year old children (4;4-5;9) participated in this study. 16 were monolingual English speakers, 19 were monolingual French speakers, and 49 were bilingual with varying exposure to French and English (16 with more exposure to English, 20 with more exposure to French, and 13 with equal exposure to both). Language exposure was determined through parent questionnaires. The bilingual children participated in a nonword repetition task and a sentence imitation task in French and English and the monolingual children participated in these tasks only for the language they spoke. The sentence imitation task came from the Recalling Sentences in Context subtest from the CELF-Preschool (2). The French NWR task came from Thordardottir et al. (2011) and the English NWR task was the CNRep by Gathercole et al. (1994). Both of the NWR tasks were scored according to Dollaghan and Campbell (1998). Study 2: 56 children with a mean age of about 5 years old split into four groups (bilingual with language impairment (LI), bilingual typically developing (TD), monolingual LI, and monolingual TD) participated in this study. The children all completed the French NWR and SI tasks from Study 1 and the Échelle de vocabulaire en images Peabody (EVIP).

Conclusions: Study 1: The association between the French NWR task and French language exposure was nonsignificant, but the English NWR task was significantly associated with English language exposure. However, the association was weaker than for SI scores and previous language exposure. Overall it appears that NWR performance is less reliant on previous language exposure. For the English NWR at the four-syllable

level the group with more exposure to French scored significantly lower than the group that had equal exposure to French and English and at the five-syllable level the group with more exposure to French scored significantly lower all other groups. For the French NWR task, there were significant differences in scores between word lengths other than between the two- and three- syllable and four- and five- syllable lengths. There were no significant differences between NWR performance in French and English. Receptive vocabulary scores were also available for these children and the English receptive vocabulary scores were significantly correlated with English NWR and SI, but not with the French measures. The same was true in reverse for French.

Study 2: On the NWR task, the two LI groups scored significantly lower than both TD groups. The LI groups did not score significantly different from each other and neither did the TD groups. The NWR task identified a difference regardless of language exposure. The SI test had a similar pattern. At a cutoff score of 82% the NWR task had a high level of sensitivity (85% and 92%) for bilingual and monolingual children respectively. It had a high level of specificity (100%) for monolingual children and a fair level of specificity (79%) for bilingual children. Any combination of two of the three measures (NWR, SI, EVIP) moved sensitivity levels to 100% for both monolingual and bilingual children and specificity levels to 100% for monolingual children. Specificity levels for bilingual children were highest with the NWR task alone. Increasing length of nonwords only significantly impacted the LI groups, except for the bilingual TD group between three and four syllables. The LI groups scored significantly lower than the TD groups at all lengths except the two-syllable length. *Relevance to current study:* This test explored differences between monolingual and bilingual children with and without

language impairment on a nonword repetition task and gave information regarding the relationship between NWR performance and language exposure. It also showed that NWR tasks can have a high level of sensitivity and specificity for identifying language impairment in children.

Windsor, J., Kohnert, K., Lobitz, K. F., & Pham, G. T. (2010). Cross-language Nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, *19*(4), 298-310. https://doi.org/10.1044/1058-0360(2010/09-0064)

Objective: To determine the relative use of Spanish and English nonword repetition tasks in identifying children with and without a language impairment. To determine if performance on a nonword repetition task is correlated across languages. *Method:* 187 children were divided into four groups: 69 in the English monolingual typically developing (TD) group, 34 in the English monolingual language impairment (LI) group, 65 in the sequential Spanish-English bilingual typically developing group, and 19 in the sequential Spanish-English bilingual language impairment group. The monolingual TD children completed the Recalling Sentences and Concepts and Directions subtests of the English version of the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4) and the bilingual children completed these subtests in English and Spanish. All children completed the Test of Nonverbal Intelligence – Third Edition (TONI-3) and a Spanish and an English nonword repetition task. The English nonword list was from Dollaghan and Campbell (1998) and the Spanish list came from Ebert et al. (2008). The tasks were scored by percent phoneme correct.

Conclusions: In the English NWR task the monolingual English typically developing group scored significantly higher overall than all the other groups. The typical

bilingual group scores overlapped with the monolingual LI group, but were significantly higher than the bilingual LI group. All groups had significantly lower performance on the longest, four-syllable nonwords and most of the differences between groups were viewed at the three- and four- syllable levels. For the Spanish NWR task the bilingual TD group scored better than all other groups. The monolingual LI group scored lower than the bilingual LI group and much lower than the monolingual TD group. The scores of the longest two syllable lengths (four and five) most clearly separated the groups. The NWR task scores were correlated for all groups except the bilingual LI group.

At a cutoff score of 78%, the English NWR task had moderate sensitivity (0.76) and specificity (0.73) for monolingual English speakers with a positive likelihood ratio of 2.78 and a negative likelihood ratio of 0.32 and high sensitivity (0.94) and low specificity (0.57) for bilingual speakers with a positive likelihood ratio of 2.20 and a negative likelihood ratio of 0.09. Only the negative likelihood ratio for the bilingual group has sufficient diagnostic accuracy. At a cutoff score of 80%, The Spanish NWR task had moderate sensitivity (0.77 and 0.58) and specificity (0.64 and 0.82) for the monolingual and bilingual groups respectively, with a positive likelihood ratio of 2.11 and a negative likelihood ratio of 0.37 for monolingual children and a positive likelihood ratio of 3.14 and a negative likelihood ratio of 0.52 for bilingual children. None of these values have sufficient diagnostic accuracy. *Relevance to current study:* This study compares the performance of monolingual and bilingual children on NWR tasks in two languages. The current study will also be testing children with NWR tasks in two languages. This study gave detailed information about sensitivity and specificity and likelihood ratios, which are important to determine the diagnostic accuracy of a nonword repetition measure.

APPENDIX B

Institutional Review Board Statement

Due to COVID-19, the current study was conducted as a secondary analysis of data that was gathered from a previous study in 2010.

APPENDIX C

Instruments

Nonword Scoring Protocol
Adapted from Gray et al., 2019

1. False starts are not errors: omit these from the placement in the nonword sequence.

du, du, tupwib—score as tupwib

w@, wl, wlft@f---score as wlft@f

2. 'Ums' and 'hu' are not errors: omit these from the placement in the nonword sequence.

Um, um, gEn, um gEnfad—score as gEnfad

Um, um, uh, wiv, um, ncktuf—score as wivncktuf

3. Line up nonword phonemes appropriately for most credit (e. g., missing syllables, phoneme addition or deletion) in the placement in the nonword sequence.

4. You may "slide" phonemes or syllables to maximize points that a child receives. Sliding a child's response to maximize points earned is acceptable under the following conditions:

a. Syllables must retain the order of the child's response:

Example:	Acceptable	Not Acceptable
Target: /nUdfegdYnyup/	n U d f e g d Y n y u p	n U d f e g d Y n y u p
Response: "metIdhu"	m e t I d h u	t I d m e h u

b. If a word can be scored in more than one way, score it in a way that the child gets the most points. For example, in the example below, you would choose to score it as Option 1 because the child receives 4 points versus 3 points (Option 2).

Example:	Option 1	Option 2
Target: /wefyUktughad/	w e f y U k t u g h a d	w e f y U k t u g h a d
Response: "yokhag"	y o k h a g	y o k h a g

c. Syllables can hang off the end or beginning of the word if lining syllables up will earn the child additional points.

Example:	Acceptable

Target: /yiktuf/ y i k t u f

Response: “tui” t u i

Example: Acceptable

Target: /kYmyeg/ k Y m y e g

Response: “dufkym” d u f k Y m

d. Syllable structure must be maintained (i.e. it is not okay to break up a CVC response structure *across* target syllables).

Example:	Acceptable	Not Acceptable
Target: /kymtup/	k Y m t u p	k Y m t u p
Response: “mup”	m u p	m u p

5. You may “pop” out syllables or extra phonemes to maximize points. Popping a syllable/phoneme out is only okay under the following conditions:

a. If the child produces more syllables/phonemes than the number of target syllables/phonemes – pop out the excess syllable/phonemes to maximize the child’s points.

Example:	Acceptable
Target: /yiktuf/	y i k t u f
Response: “yikatuf”	y i k(a)t u f

(note that in this case that, although the child got all of the phonemes correct, the “whole word” would not be counted as correct because of the extra syllable.

Example:	Acceptable
Target: /yitgYm/	y i t g Y m
Response: “yisgrYm”	y i s g(r)Y m

6. Line up vowels with vowels and consonants with consonants.

7. Line up the child’s response from the beginning of the nonword, unless there is a phoneme similar to the one in the nonword (e.g. /b/ for /p/), in which case you may slide it over as it is the closest substitution.

Example:

Target: /wefyUktughad/	w e f y U k t u g h a d
Response: “dug”	d u g

English non words
Subject ID

Audiofile start point

Transcriber
Examiner

	Target	c	v	c	v	c	v	c	v	c
1)	neɪb	n	eɪ	b						
2)	voʊp	v	oʊ	p						
3)	taʊdʒ	t	aʊ	dʒ						
4)	doɪf	d	oɪ	f						
5)	teɪvak	t	eɪ	v	a	k				
6)	tʃaʊvæg	tʃ	aʊ	v	æ	g				
7)	vætʃaɪp	v	æ	tʃ	aɪ	p				
8)	noɪtʃaʊf	n	oɪ	tʃ	aʊ	f				
9)	tʃinoɪtaʊb	tʃ	i	n	oɪ	t	aʊ	b		
10)	naɪtʃaveɪb	n	aɪ	tʃ	a	v	eɪ	b		
11)	doɪtaʊvæb	d	oɪ	t	aʊ	v	æ	b		
12)	teɪvoɪtʃaɪg	t	eɪ	v	oɪ	tʃ	aɪ	g		
13)	veɪtatʃaɪdoɪp	v	eɪ	t	a	tʃ	aɪ	d	oɪ	p
14)	dævoʊnoɪtʃɪg	d	æ	v	oʊ	n	oɪ	tʃ	i	g
15)	naɪtʃoɪtaʊvub	n	aɪ	tʃ	oɪ	t	aʊ	v	u	b
16)	tævatʃinaɪg	t	æ	v	a	tʃ	i	n	aɪ	g

Spanish non words
Subject ID

Audiofile start point

Transcriber
Examiner

	Target	c	v	c	c	v	c	c	v	c	c	v	c	c	v	c
1)	gauβeɾ	g	au		β	e	ɾ									
2)	meɾɣui	m	e	ɾ	ɣ	ui										
3)	tiŋkua	t	i	ŋ	k	ua										
4)	meɾfas	m	e	ɾ	f	a	s									
5)	duɾbiepos	d	u	ɾ	b	ie		p	o	s						
6)	tʃeɾuɣua	tʃ	e		ɾ	u		ɣ	ua							
7)	tintʃauβel	t	i	n	tʃ	au		β	e	l						
8)	ɾutʃetua	ɾ	u		tʃ	e		t	ua							
9)	kitʃeɾupia	k	i		tʃ	e		ɾ	u		p	ia				
10)	meɾantʃutin	m	e		ɾ	a	n	tʃ	u		t	i	n			
11)	maʊteipoitin	m	aʊ		t	ei		p	oi		t	i	n			
12)	fiɾutʃepia	f	i		ɾ	u		tʃ	e		p	ia				
13)	hustinɾunaitʃe	h	u	s	t	i	n	ɾ	u		n	ai		tʃ	e	

Spanish non words
Subject ID

Audiofile start point

Transcriber
Examiner

14)	biŋkuamiɛferɣui	b	i	ŋ	k	ua		m	iɛ		f	e	ɾ	ɣ	ui	
15)	nueβurdaiɣuipos	n	ue		β	u	ɾ	d	ai		ɣ	ui		p	o	s
16)	pentʃufaitinβel	p	e	n	tʃ	u		f	ai		t	i	n	β	e	l