Bilingual Children's Performance on Three Nonword Repetition Tasks: The Role of Language Experience and Ability

By

Simone Hüls
M.A., University of Alabama in Huntsville, 2004
B.A., University of Alabama in Huntsville, 2001

Submitted to the graduate degree program in Speech-Language-Hearing: Sciences & Disorders and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

____

Chair: Marc Fey, PhD, CCC-SLP

____

Co-Chair: Holly Storkel, PhD, CCC-SLP

____

Betty Bunce, PhD, CCC-SLP

____

Alison Gabriele, PhD

____

Ana Paula G. Mumy, MS, CCC-SLP

Date Defended: 10 May, 2017

The dissertation committee for Simone Hüls certifies that this is the approved version of the following dissertation:

Bilingual Children's Performance on Three Nonword Repetition Tasks: The Role of Language Experience and Ability

Co-Chair: Marc Fey, PhD, CCC-SLP

Co-Chair: Holly Storkel, PhD, CCC-SLP

Date Approved: 10 May, 2017

Abstract

Nonword repetition (NWR) tasks represent one assessment tool for Specific Language Impairment (SLI). The use of such tasks has been established and verified for monolingual children. However, the diagnostic accuracy of NWR tasks for bilingual children has had variable results and must address several unique characteristics of this population. Gaps in research relate to which task characteristics influence task performance. The purpose of the current study was to explore children's performance on three NWR tasks that differ in design: a norm-referenced task based in English (*Comprehensive Test of Phonological Processing, Second Edition; CTOPP-2),* a "quasi-universal" task based in the sounds most commonly found in the world's languages (QU), and a "language neutral" (LN) task tailored to Spanish and English speech sounds. The children's performance on the three tasks was examined in light of their time spent listening to and speaking English and language ability in Spanish and English, as measured by the *Bilingual English Spanish Assessment* (BESA). Finally, the different versions of the NWR task were compared in terms of which task is most predictive of a performance on the BESA that is indicative of language impairment. Results indicated that the LN task and the CTOPP-2 task performance may have been influenced by a child's language experience in English. All three tasks were significantly related to the participants' language ability, as measured on the BESA. Finally, analyzing the data using quantile regression indicated that for all three tasks, the lower quantiles of the BESA Index score (i.e., participants' language ability) was more strongly related to the performance on the NWR tasks than the BESA performance at higher quantiles. All three tasks hold promise for diagnosing Spanish-English bilingual children with SLI, with the QU task performance showing more independence in terms of language experience. In conclusion, while

LN and the CTOPP tasks revealed participants' performance based on language experience, the QU task is the most promising task for an assessment tool for bilingual children.

Keywords: language impairment, bilingual, multilingual, assessment, nonword repetition

Acknowledgments

This project, like most research, would not have been possible without the help of a large research community. I feel incredibly lucky to have ended up at the University of Kansas, with a group of students and mentors who are dedicated to furthering their area of study, engaged in high-quality research and building an academic community. To my advisors, Marc Fey and Holly Storkel, thank you for teaching me, guiding me, and showing me what it takes to be a part of a research community. You have both been incredible mentors, and I am thankful that fate placed me in your path. To the other member of my committee, Alison Gabriele, Betty Bunce, and Ana Paula Mumy, thank you for feedback and guidance as I completed this project. To my fellow "lab mates:" thank you for the engaging discussions, sharing your own research efforts with me, and the countless ways in which you have all helped me complete this study. I especially want to thank Rouzana Komesidou for our conversations on international matters and US culture; I don't think I will ever have such an intelligent and funny co-worker again. Although I am sure you will travel the world, I hope that our paths continue to cross. To Veronica Fierro: there are not enough ways that I can thank you for the hours that you spent helping me complete this study. From translating forms, administering tests, and providing information on Spanish language acquisition, you have been involved in all aspects of my dissertation. For that, I feel like you should be listed as a co-author, so know that your name is there in spirit. Thank you to all of the students who were involved in helping me complete my research. I cannot list you all, but special thanks to Nicole Schuh, who was there from the inception of the idea to the completion of the final project.

I could not have completed this dissertation or this degree without the help of my family. I would like to thank my parents for instilling in me a love of learning and a curiosity about how the world works. More importantly, they have taught me compassion and caring, which are the foundations of my research interest in helping those that are not receiving services they need. You have endowed me with critical thinking skills and the pursuit of equality and fairness, always encouraging me to strive to achieve. Without your support I would have never pursued this degree. Finally, to my wife Stacy and my daughter Izzy: thank you for your encouragement, patience, and support. Stacy, you led the way in our pursuits to get a PhD, and I am so happy to follow in your footsteps. Thank you for supporting me always, for your love and for cheering me on, providing support and for pushing me when I needed a nudge. You always seem to know what I need, and your free psychological counseling has helped make the dissertation process relatively painless. I am so lucky to have you as my partner and I look forward to the next phase in our adventure together. To Izzy, thank you for putting up with the incredibly long time that I had to work on my "Doktorarbeit." Thank you for your encouragement and thank you for attempting to recruit participants for me at your preschool. You are already a scientist and have a curiosity that cannot be eclipsed. I love you both and could not have completed my dissertation without your love and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter1: Introduction

Specific Language Impairment (SLI)[1] is a developmental language disorder characterized by low performance on language measures, without concomitant deficits in other areas such as cognitive, neurological, or physical ability. This disorder affects approximately 7% of children and is therefore of clinical relevance for school-based Speech Language Pathologists (Tomblin, Records, & Zhang, 1996). Children with SLI are disadvantaged in their daily communication, which may cause social-emotional difficulties in addition to academic problems such as difficulty learning new vocabulary (Leonard, 2014). In addition, children with SLI are at greater risk of developing reading disorders and demonstrate communication deficits into their adult years (Catts, Fey, Tomblin, & Zhang, 2002; Poll, Betz, & Miller, 2010). Therefore, early identification of SLI is of utmost importance in ameliorating the long-term impact of the disorder. However, because SLI is a behaviorally classified disorder, diagnosis is often based on a multifaceted assessment; many clinicians and researchers argue that there is no "gold standard" assessment for SLI (Tomblin et al., 1996).

In their epidemiological study, Tomblin, Smith, and Zhang (1997) estimated that approximately 70% of 5-year-old children with SLI are not accurately identified and therefore do not receive services. It can be assumed from the previously discussed information regarding monolingual children with SLI that bilingual children with SLI are at an even greater risk of under- and over-diagnosis due to the heterogeneity that is inherent in this population. It is

---

[1] Several researchers have made a case for the term "Primary Language Impairment" to more accurately denote this disorder (Kohnert, Windsor, & Ebert, 2009). However, this paper will use the term more commonly found in the literature. For a discussion of the merits of different terminologies, see Bishop (2014).

assumed that bilingual children exhibit a similar prevalence rate, although an epidemiological study such as that completed by Tomblin et al. (1997) has yet to be completed with bilingual children (Gillam, Peña, Bedore, Bohman, & Mendez-Perez, 2013). Bilingual children may vary on any number of factors related to their language exposure, such as the age of first exposure to the second language, their language experience in terms of the length, amount, and type of exposure, and syntactic or phonological characteristics of the languages known by the child. Knowledge-based assessments such as vocabulary or grammar measures are heavily influenced by such background characteristics, making them unsuitable for identifying SLI in bilingual children. Thus, it is difficult to distinguish a typical child who underperforms because of lack of experience in English from peers with SLI who underperform due to language impairment.

One possible solution to the problematic nature of knowledge-based assessments is to use a processing-based measure such as a nonword repetition (NWR) task. These tasks have been explored as a clinical tool for use with monolingual children since the 1980s. NWR tasks require a child to immediately repeat a phonologically unfamiliar nonword, which carries no semantic information. Measures that are based on prior knowledge, such as vocabulary tests, are influenced by child-external factors such as socioeconomic factors (SES), making accurate assessment of SLI difficult (Hoff, 2003). Because NWR tasks are processing- rather than knowledge-based, they have been shown to be less influenced by language experience or factors such as a child's SES, making them appropriate for use with linguistically diverse populations (Campbell, Dollaghan, Needleman, & Janosky, 1997; Engel, Santos, & Gathercole, 2008; Roy & Chiat, 2004). While NWR tasks are less influenced by language experience, they are not completely independent of it. Research has shown that bilingual children's performance on these tasks is influenced by language experience and vocabulary knowledge in the language on which the task is based (Ebert, Pham, & Kohnert, 2014; Lee & Gorman, 2012; Parra, Hoff, & Core, 2011; Peña, Bedore, & Zlatic-Giunta, 2002; Summers, Bohman, Gillam, Peña, & Bedore, 2010; Thorn & Gathercole, 1999). Thus, a child's unique language experience profile may impact task performance and thus the diagnostic accuracy of a particular NWR task.

One method of minimizing the role of language experience in task performance is to use the "quasi-universal" (QU) nonword repetition task, which uses only the sounds that are common to most of the world's languages (Chiat, 2015). This task potentially reduces the statistical confound of language experience and is possibly more accurate in screening for, and identifying those bilingual children who have SLI. However, research on the diagnostic accuracy of the QU task is still in its initial stages. Only one published study has examined the diagnostic accuracy of the task for Dutch-speaking bilingual children from various language backgrounds. This study determined that the task has good diagnostic accuracy for this population (Boerma et al., 2015). Another possible solution for the problem of language confounds is to create a task that is "language neutral" (LN) or tailored to a child's specific language combination. Such a task would use only the sounds and segments that are found in both of a child's languages, therefore eliminating the impact of prior language experience. No prior research has been conducted regarding the diagnostic accuracy of such a task.

The purpose of the following study was to compare bilingual children's performance on a "quasi-universal" task, a "language neutral" task, and a norm-referenced task based in English, the *Comprehensive Test of Phonological Processing-2nd Edition (CTOPP-2).* First, the research examined how a child's performance on the three tasks changes with varying levels of exposure to English, quantified as percentage of time spent speaking/listening to English. Second, the study examined how a child's performance on each task changed with varying levels of language ability in Spanish and English, as measured by performance on the *Bilingual English Spanish Assessment* (BESA). Finally, the current study explored the relationship between children's performance on each of the three tasks in identifying bilingual Spanish-English speaking children with low language ability.

This study's significance is both theoretical and applied. By manipulating the NWR task design, the current study adds to the research regarding bilingual children's performance on NWR tasks and these tasks' ability to discriminate bilingual children with and without SLI. The ultimate goal of examining children's performance on different NWR tasks is to determine which task has the most potential for use by a monolingual Speech-Language-Pathologist to accurately assess a bilingual child using English-based or language-neutral assessments.

The remainder of this document is organized in the following way: Chapter Two discusses the conceptual framework of this study. It outlines the characteristics and language profiles of monolingual and bilingual children with SLI, provides relevant information regarding the challenges of assessing bilingual children for SLI, describes the use of NWR tasks when assessing monolingual children for SLI, and examines the use of such tasks in the assessment of bilingual children with SLI. Factors related to language experience and vocabulary, as well as the characteristics of different language combinations, are discussed. Chapter Three focuses on the study's method. The overall research approach, population and sampling techniques, instruments used, and procedures completed are discussed. Chapter Four provides the statistical results, and Chapter Five offers a summary, conclusions, and clinical implications.

Chapter 2: Conceptual Framework

Children with Specific Language Impairment (SLI) exhibit deficits in grammar, vocabulary, and phonological processing in the absence of other deficits such as a hearing impairment, low nonverbal IQ, or neurological damage (Leonard, 2014). It is estimated that 7.8% of Kindergarten-age children exhibit SLI, a rate that indicates that this is one of the most common language disorders for children (Tomblin et al., 1997). Children with SLI are usually impaired in phonology (i.e., phonological awareness or phonological memory), morphology, and syntax. In English, children with SLI typically exhibit several grammatical deficits, such as errors in finiteness marking like the third person singular *–s* or the past tense *–ed* (Leonard, 2014). In other languages, the grammatical markers differ, and may be exhibited in diverse ways, such as through errors in clitics (Dispaldro, Leonard, & Deevy, 2013; Jakubowicz, Nash, Rigaut, & Gerard, 1998; Theodorou & Grohmann, 2015) or word order (Conti-Ramsden & Windfuhr, 2002; Hansson, Nettelbladt, & Leonard, 2000). Therefore, although the underlying deficit is theorized to be the same, the behavioral expression of the deficit will vary from language to language.

SLI impacts a child's communication ability, academic achievement, and reading development, the effects of which continue into adulthood (Catts et al., 2002; Leonard, 2014). Early identification and intervention may ameliorate the deficits of the disorder, increasing the child's academic achievement and reducing the resources required to aid the child's development. Because SLI is defined through a pattern of behavior, diagnosis of SLI usually results from a multifaceted assessment by a multidisciplinary team. In other words, SLI is often not diagnosed using a single norm-referenced measure (Leonard, 2014). The difficulty of accurately diagnosing SLI in the early elementary grades results in a larger rate of under-diagnosis for monolingual children. In their epidemiological study, Tomblin et al. (1997) determined that as many as 70% of kindergarten-age children with SLI are not identified and therefore not receiving services. One can suppose that this rate of under-identification would be even greater in bilingual or multilingual children (Gillam et al., 2013). In their study, Grimm and Schulz (2014) determined that under-diagnosis was more common than over-diagnosis in bilingual children. Thus, while already a complex process for monolingual children, accurate assessment of SLI in bilingual children is a multifaceted interplay of factors that must be explored through systematic research.

Assessment of SLI in Bilingual Children

The assessment of SLI in bilingual children is problematic because under- and over-diagnosis can have an immediate economic and educational impact. In addition, bilingual children encompass a large percentage of the total student population. For example, in Kansas, 6.0 to 9.9 percent of school-aged children are enrolled in English Language Learning services, a rate that will likely increase over time (nces.ed.gov). These children are enrolled in school districts across the state, and are not limited to urban areas. Thus, it is highly likely that Speech-Language Pathologists in a range of educational settings will encounter bilingual children in need of language assessments. While the American Speech-Language-Hearing Association recommends that a bilingual child should be assessed in both of their languages, Speech-Language Pathologists who are attempting to diagnose a bilingual child are not likely to speak the child's first language (American Speech-Language-Hearing Association, 2014). In a national survey of clinicians conducted in 2013, only 5% of clinicians met the criteria for being bilingual. The assessment of children in all of their languages is also often not possible for practical reasons because assessments in the child's first language may not exist, be difficult to procure, or be normed on monolingual rather than bilingual children. In addition, qualified interpreters may be difficult to recruit (Grimm & Schulz, 2014).

The assessment of bilingual children is further complicated by the heterogeneity of this population, which may be characterized by child external factors (e.g., socioeconomic status, length of exposure, percentage of English use at home, or the "richness" of the English environment) and child internal factors (e.g., chronological age, first language background, or language learning aptitude) (Paradis, 2015). For example, determining whether a second language learner is sequential or simultaneous greatly impacts the type of assessment and interpretation of assessment results, as these children have divergent patterns of language acquisition. Any language assessment conducted in English will be heavily influenced by the child's experience with the language, thus making it difficult to separate a language difference from a language disorder. Children who are acquiring English sequentially will differ from typically developing English monolingual children; however, as their exposure to English increases they may also differ significantly from children learning their first language. The relationship between the length of exposure, current amount of exposure, and language knowledge is therefore in constant fluctuation. Thus, knowledge-based measures such as vocabulary and grammatical assessments have only limited utility in identifying bilingual children with a language disorder, especially when administered in only one of the child's languages. Even when administration of tests in both of the child's languages is possible, the fact that the assessments were likely normed on monolingual speakers of the language limits their application to bilingual children, as these children will likely differ in their performance from monolingual speakers. The influence of current language exposure on test results will likely fluctuate over time. For example, a child may have limited exposure to English during the summer months but nearly equal exposure to both languages during the school year. Current exposure to each language will likely influence vocabulary knowledge, whereas cumulative exposure may be related to syntax development. These factors must be considered when testing bilingual children.

Bilingual Assessment Measures. Because assessment of bilingual children must consider the various factors that influence language development in this population, it is possible to utilize alternate assessment methods appropriate for bilingual children. These include parent interviews/questionnaires and processing-dependent measures. Processing-dependent measures attempt to eliminate the influence of prior language experience, ensuring a more accurate assessment of the "difference versus disorder" dilemma. Interviews or questionnaires involve gathering background information from parents, teachers, and other caregivers that determine whether a child is at risk of having a language disorder (Paradis, Schneider, & Duncan, 2013; Restrepo, 1998). This method often requires the use of an interpreter, making it less feasible for many Speech-Language Pathologists than assessments administered in English.

One of the major purposes of a parent interview or questionnaire is for a monolingual clinician to gather information on a child's first language development or risk factors for language impairment. If a bilingual child has a language disorder, language deficits will be apparent in both of the child's languages (Paradis, Genesee, & Crago, 2011). Thus, information on first language (L1) performance is essential in interpreting any findings based on the child's second language (L2) performance. For example, it is possible that a school-age child has less exposure to their L1 than the L2 because they are spending most of their time in the English-language school environment. Thus, a child who scores low in the L2 will likely also score low on language testing of the L1, indicating that the child has SLI. If, on the other hand, a child receives more exposure to their L1 than L2, a low score on a language assessment in the L2 may merely reflect a lack of experience in this language. However, SLI cannot be ruled out until testing is completed in the L1. If this child scores within typical limits on the L1 testing, this would indicate a language difference rather than language impairment. A parent questionnaire can provide crucial information on L1 development when testing in the L1 is not possible. Thus, although the questionnaire/interview will not lead to a diagnosis independent of direct assessments of the child's abilities, background information gleaned from alternate assessments may play a role in confirming or disconfirming a diagnosis of SLI. In fact, questionnaires such as the Spanish Ages and Stages Questionnaire (Bricker et al., 1999), the Spanish vocabulary inventory (Inventarios del Desarrollo de Habilidades Comunicativas Palabras y Enunciados) (Jackson-Maldonado, Bates, & Thal, 1992), the Preschool Language Scale, Spanish Edition (Zimmerman, Steiner, & Pond, 2012), and parent report of the child's three longest utterances all illustrated good concurrent validity with each other and reliably identified toddlers with a language delay (Guiberson, Rodríguez, & Dale, 2011). In other words, parent reports of vocabulary knowledge in the L1 has been shown to be a valid form of assessment, which has good concurrent validity with standardized measures of vocabulary knowledge (Mancilla-Martinez, Gámez, Vagh, & Lesaux, 2015).

Other areas assessed through parent interviews/questionnaires include the child's current language ability, family history of language difficulties, and current language input and use. Detailed parent questionnaires that include this information have shown good validity in determining which children are at risk for SLI (Paradis, Emmerzael, & Duncan, 2010) and for reporting the amount of L1 input, which in turn predicts language ability in the L1 (Gutiérrez-Clellen & Kreiter, 2003). Parents are, understandably, less reliable in reporting a child's L2 input since they are likely not physically present in the child's L2 environment. Similarly, a child's teacher is less reliable in estimating a child's L1 input (Gutiérrez-Clellen & Kreiter, 2003). To date, no study known to this author has directly assessed whether parent report of the amount of L1 and L2 input correlates to the actual amount of input processed by the child. However, parent ratings of different use and proficiency in both languages correlated with observed outcomes (Bedore et al., 2011; Gutiérrez-Clellen & Simon-Cereijido, 2010; Kohnert, Windsor, & Yim, 2006; Summers et al., 2010). Despite the potential drawback of inaccuracy, parent questionnaires provide critical information to a monolingual Speech-Language Pathologist (SLP) during the assessment process. In addition to questionnaires, another alternative assessment for bilingual children is the use of processing-based measures, such as nonword repetition tasks. These tasks address the limitations of knowledge-based assessments, such as vocabulary or grammar measures in the child's L2. One processing-based measure, nonword repetition, is discussed in more detail below.

When examining the language ability for Spanish-English bilingual children, it is possible to use a relatively new standardized assessment, the *Bilingual English Spanish Assessment* (BESA). This assessment has been normed on bilingual speakers, giving it the unique status of being a valid assessment of language disorders in bilingual children (Peña, Gutiérrez-Clellen, Iglesias, Goldstein, & Bedore, 2014). The BESA, which assesses a child's morphosyntax, semantics, phonology, and pragmatics, was used as the standardized measure for assessing a child's language ability in this study. While this assessment is a promising assessment tool, it requires an SLP to be bilingual in Spanish and English. Further information on the BESA will be provided in the method section.

Nonword Repetition Tasks in Monolingual Children

Research on processing-based assessment measures includes a focus on nonword repetition (NWR) tasks. Much of the research on NWR tasks has focused on monolingual English-speaking children. Research shows a robust and well-documented lower performance of monolingual children with SLI on NWR tasks (Archibald, 2008; Estes, Evans, & Else-Quest, 2007). Thus, these tasks are one established method of identifying monolingual children with SLI and differentiating them from their typically developing peers (Bishop, North, & Donlan, 1996; Conti-Ramsden, Botting, & Faragher, 2001).

In their meta-analysis, Estes et al. (2007) examined four different NWR tasks across 23 studies. The authors identified several factors that impacted the differences in scores between diagnostic groups: a higher proportion of multisyllabic nonwords, inclusion of consonant clusters, and all-or-nothing scoring. In addition, longer items on an NWR task reveal larger differences between the performance of typically developing (TD) children and those with SLI (Archibald & Gathercole, 2006; C. Dollaghan, Campbell, Needleman, & Dunlosky, 1997; Gathercole & Baddeley, 1990). Therefore, a NWR task with longer, more complex items that is scored in an all-or-nothing manner leads to larger difference in accuracy of TD monolingual children and those with SLI. The diagnostic accuracy of different NWR tasks has been established in the research. For example, the Children's Test of Nonword Repetition (Gathercole & Baddeley, 1990), has a reported sensitivity (the number of true positives divided by the total number of individuals with the condition) of 78% and a specificity (the number of true negatives divided by the total number of individuals without the condition) of 87%, using a cutoff of the 16^{th} percentile, indicating that this task has clinical utility in identifying monolingual children with SLI (Conti-Ramsden et al., 2001). Overall, the data reviewed by Estes et al. (2007)

indicated that monolingual children with SLI performed 1.27 standard deviations below children without the disorder. While there was considerable variation in the studies reviewed, the results indicated that NWR tasks are a reliable means of identifying monolingual children with SLI. In almost all of the reviewed studies, regardless of the task used, there was a statistically significant difference between children with and without SLI. Therefore, NWR tasks are a clinical assessment tool that may be used in combination with other tasks to accurately identify monolingual children with SLI.

An additional aspect of NWR tasks explored in the literature is that they are less influenced by factors such as socio-economic status, which has consistently been shown to affect knowledge-based measures such as vocabulary tests (Hoff, 2003; Rowe, 2008; Roy & Chiat, 2013). While knowledge-based measures show differences between high- and low-SES groups, performance on NWR tasks is equal between these groups, eliminating the impact of SES on task performance (Campbell, Dollaghan, Needleham, and Janosky, 1997).

Key Challenges to Using Nonword Repetition Tasks in Bilingual Children

Research on NWR tasks for the assessment of bilingual children is limited and presents great variability in diagnostic accuracy. This variability likely stems from English-language learner characteristics, task characteristics, and elements of research design. Therefore, while NWR tasks are superior to knowledge-based tasks such as vocabulary measures, they are not a panacea for the problem of how to accurately assess bilingual children for SLI. While these tasks do not directly draw on vocabulary, morphology, or syntax, they are still influenced by such language characteristics. The question remains of how to eliminate the influence of the child's language experience, vocabulary knowledge, or age of first exposure, as these factors decrease diagnostic accuracy of tasks.

One possible solution is the use of a “quasi-universal” (QU) nonword repetition task such as that proposed by Chiat (2015). This type of task, discussed in detail in the method section, utilizes only sounds that are common in the world’s languages, thus potentially minimizing the influence of prior language experience. As mentioned earlier, research on the diagnostic accuracy of the QU task is still in its initial stages but shows promising initial results. In the first published study on bilingual speakers of Dutch, Boerma et al. (2015) determined a positive likelihood ratio of 12.5 and a negative likelihood ratio of .18, indicating good discriminability for the QU task. Research on English-based NWR tasks has shown that items that are prosodically structured, are longer, have more complex segmental structures, and have lower frequency phoneme sequences more accurately distinguished bilingual children with SLI from their typically developing peers (Chiat, 2015). The relevant factors involved in using NWR tasks with bilingual children will be discussed in detail below.

Language experience/exposure. In bilingual children, language experience plays a large role in vocabulary knowledge, which in turn impacts NWR performance (Peña et al., 2002). For example, Parra et al. (2011) examined Spanish-English bilingual children’s performance on a NWR task that included 12 English-like nonwords and 12 Spanish-like nonwords and found that the percentage of English language exposure was positively correlated with English NWR performance. In addition, Spanish language exposure correlated with Spanish NWR performance in these participants. Similarly, Ebert et al. (2014) found that correlations between NWR performance in English and Spanish and language exposure indicated that NWR tasks assess some language-specific aspects of processing. Summers et al.'s (2010) study involving Spanish-English bilingual children also found that NWR performance was significantly correlated with cumulative language exposure in these children. Therefore, the research provides support for the impact of language experience on NWR performance. One complicating factor for bilingual children is that their relative exposure to the first and second language (L1 and L2) changes over time, with language dominance often shifting to the L2 during the school years. This shift often results in a pattern of "mixed dominance," whereas children's linguistic profiles may include semantic strength in one language and stronger morpho-syntactic skills in the other language (Paradis, 2015). Therefore, the relationship between language experience and NWR performance is not static over time.

Vocabulary knowledge and NWR tasks. In addition to measuring the effects of language exposure, researchers have examined vocabulary knowledge and its effect on NWR performance. For example, Thorn and Gathercole (1999) found that greater vocabulary knowledge leads to higher levels of recall of nonwords in both of a bilingual child's languages. In addition, Lee and Gorman (2012) examined the effect of bilingual children's vocabulary knowledge on English NWR performance. The participants included children with various language backgrounds: monolingual English speakers, Korean-English bilingual children, Chinese-English bilingual children, and Spanish-English bilingual children. Vocabulary knowledge and NWR performance was related in all groups. The authors concluded that bilingual individuals do not necessarily do worse on English NWR tasks than L1-based tasks, but that their performance in each task is based on vocabulary knowledge in the language on which the task is based. Finally, Paradis (2015) determined that performance on the CTOPP, a standardized measure normed on monolingual English-speaking children, is strongly influenced by receptive vocabulary size in English, as well as the phonological characteristics of the child's first language.

NWR performance and age. The relationship between vocabulary knowledge and NWR performance may change with age. For example, Lee, Kim, and Yim (2013) examined monolingual Korean and bilingual Korean-English toddlers' performance on a Korean-based NWR task. The authors compared the children's performance on the task and examined the relationship between the NWR task and vocabulary measures. While the two groups differed significantly on their vocabulary measures, they showed no difference in their performance on the NWR task. In another study involving young children, Brandeker and Thordardottir (2015), examined NWR performance in English and French bilingual toddlers. The authors found moderate to strong associations between language exposure and vocabulary. However, the relationship between language exposure and NWR performance was weak. The authors concluded that language exposure affects vocabulary knowledge more than NWR performance, at least in young children. Reviews such as Coady and Evans (2008) have shown a shift around age five in terms of what language processes are involved in NWR performance. Therefore, research with children younger than five should be interpreted with caution when examining the impact of vocabulary knowledge on NWR performance.

Impact of phonological characteristics. As mentioned earlier, the difficulty of assessing the role of language experience for bilingual children stems from the interaction between language experience and vocabulary. In addition, language experience impacts phonological memory and awareness (Summers et al., 2010). NWR tasks measure phonological short-term memory but are also influenced by long-term lexical knowledge. For example, bilingual NWR accuracy is greater for wordlike items than non-wordlike items, especially for younger students (Munson, Kurtz, & Windsor, 2005). This is because nonwords with phoneme sequences that are common in the L1 or L2 are more accurately repeated since children partially rely on long-term lexical storage during the task. According to several researchers, wordlike nonwords activate phonological neighbors in the L1 or L2, thus facilitating recall (Edwards, Beckman, & Munson, 2004; Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997). Bilingual children, because they have less input in each single language, often have weaker underlying phonological representations of their second language (Engel de Abreu, Gathercole, & Martin, 2011; Parra et al., 2011). Because of this, NWR task performance improves when the nonwords are constructed using phonological repertoires and phonotactic rules of the native language of the child, assuming that this is the dominant language (Masoura & Gathercole, 1999; Thorn & Gathercole, 1999). For example, Masoura and Gathercole (1999) found that Greek bilingual children, who were language dominant in Greek, were more accurate in a Greek NWR task than an English one. Similarly, Thorn and Gathercole (1999) found that English monolingual children scored worse on a French-based task than bilingual children, who scored similar on the French task and an English-based task. Summers et al. (2010) found that Spanish-English bilingual children produced Spanish-based nonwords more accurately than English-based nonwords. Overall, short-term memory is better in the native language than in the L2, as long as the native language is the dominant language.

Another factor to consider in using NWR tasks with bilingual children is the phonological overlap between the child's languages. Parra et al. (2011) posited that the shared variance in performance between their English- and Spanish-based NWR tasks could be due to the phonological overlap between Spanish and English, indicating that NWR performance may differ according to the phonological similarity between the child's L1 and L2. Thus, languages with less phonological overlap would lead to more independent NWR scores in each of the languages. Interestingly, in Ebert, Kalanek, Cordero, and Kohnert's (2008) study, Spanish NWR performance was better than English, although the children in this study had a clear dominance in English. The researchers posited that this pattern might be due to the characteristics of the Spanish NWR task items, which were longer but had a simpler syllable structure, a theory that mirrors that of Parra et al. (2011), who posited that phonological overlap between the L1 and L2 influenced performance.

For the current study, phonological characteristics of Spanish and English were considered. Table 1 outlines the phoneme inventories and phonological characteristics of Spanish and English, indicating a sizeable phonological overlap between the two languages. Despite this overlap, several English phonemes do not exist in Spanish (ŋ, dʒ, v, θ, ð, z, ʒ, h, ɹ, j) and several Spanish phonemes do not exist in English (ɲ, ɟ, ʎ, x). In terms of vowels, Spanish has a much simpler vowel system than English, with ten vowels compared to English's 19 vowels. In terms of segmental structure, both languages include clusters and various consonant-vowel structures, but English represents the more complex system overall. A CV combination is most common in both of the languages. In addition, when produced in unstressed syllables, English vowels are reduced whereas Spanish vowels are not. In terms of the stress pattern of the two languages, Spanish is a syllable-timed language, with primary stress occurring on the penultimate syllable approximately 80% of the time (McLeod, 2007). Thus, the relatively simpler phonological characteristics of Spanish could lead to higher task performance, regardless of a child's language dominance.

Finally, a child's typical phonological development in each of their languages could impact task performance. Table 2 summarizes the developmental norms for monolingual children who speak either Spanish or English. Research on the phonological development of bilingual children is somewhat limited, thus restricting the interpretation of phonological development's impact. Studies of 3-year-old bilingual Spanish-English speaking children illustrated a lower intelligibility rating, more consonant and vowel errors, and more uncommon error patterns than monolingual children of either language (Gildersleeve, Davis, & Stubbe, 1996; Gildersleeve-Neumann & Davis, 1998). However, by the age of four, these differences in phonological development largely disappear and typically developing bilingual children performed commensurate with their monolingual peers (Goldstein & Washington, 2001; Goldstein, Fabiano, & Washington, 2005). As can be seen from Table 2, Spanish does not have any phonemes that are acquired after the age of 6;0, whereas English has several such phonemes.

In conclusion, the use of NWR tasks with bilingual children necessitates consideration of several factors, including the language experience and vocabulary knowledge of the children as well as the phonological features of the two languages. The goal of any NWR tasks developed for use with bilingual children should be to maximize the differences between typically developing children and children with SLI, while controlling for factors such as language experience, vocabulary knowledge, and L1/L2 phonological characteristics, minimizing the differences between typically developing bilingual children varying in their L1/L2 language experience (Chiat, 2015).

One newly proposed assessment method for bilingual children is using a "quasi-universal" nonword repetition task. This task minimizes the effect of language experience, vocabulary knowledge, and phonological features by using a limited number of consonants, most of which are early-acquired sounds, and only three tense vowels. The utilized consonants and vowels are common to most of the world's languages. This task, developed by Chiat (2015), includes 16 items, from 2-5 syllables in length. The current study utilized a recording of task items that placed equal stress for all of the task syllables. Task items were produced using the phonetic realizations of English. The task thus was not tailored toward a specific combination of L1 and L2 but rather minimized the impact of language experience and phonological differences by using the most common sounds and prosodic structure. This task has the potential to address some of the limitations of NWR tasks based in a particular language and has been shown to be a more accurate diagnostic tool than a language-specific task (Boerma et al., 2015).

Rather than making a “universal” task, another possibility is the development of a “language neutral” (LN) task that is tailored to a specific combination of languages. Such a task would eliminate all phonemes that are not common to the languages spoken by the child. In addition, such a task would not include any real words in either of the child’s languages and would represent the phonotactic characteristics of both of the child’s languages. While an initial pilot study has been conducted with Arabic-English bilingual adults (Huls, Storkel, and Schuh, 2015), no research has been conducted regarding the diagnostic accuracy of such a task.

In summary, the information presented thus far discussed NWR performance in monolingual children, the use of NWR with bilingual children, and possible tasks that address limitations of past research on the use of NWR tasks with bilingual children. The following section will outline previous research findings regarding diagnostic accuracy of tasks based in a specific language (English, Spanish, French, Dutch, and Icelandic).

Diagnostic Accuracy of NWR Tasks Used in the Assessment of Bilingual Children

Diagnostic accuracy metrics. Various methods may be employed to determine the accuracy of a diagnostic tool, such as the area under the receiver operating characteristic (ROC) curve, calculating sensitivity and specificity, and determining the positive and negative likelihood ratios of a test (Dollaghan & Horner, 2011). Dollaghan and Horner (2011) make the case that likelihood ratios are the most informative measure of language screening accuracy.

Therefore, the review will outline the likelihood ratios found in studies related to the diagnostic accuracy of NWR tasks in the use of assessing bilingual children.

The positive likelihood ratio is calculated by dividing the sensitivity of the measure by *1-specificity*. This measure represents the extent to which a positive test result increases the likelihood that a child has SLI. The negative likelihood ratio, on the other hand, measures the extent to which a negative test decreases the likelihood that a child has SLI. This measure is calculated by dividing *1-sensitivity* by the specificity. In order to determine sensitivity, specificity, and likelihood ratios for the tasks employed, researchers need to determine cutoff scores at which to assign children to a diagnostic category (typically developing or having SLI). Most researchers reported a cutoff score around 70% percent consonants correct on the NWR task but cutoffs varied across studies. The cutoff scores reported in the literature on the diagnostic accuracy of NWR tasks for assessing bilingual children range from 33% (Girbau & Schwartz, 2008) to 93% (Kohnert, Windsor, & Yim, 2006). The rule of thumb described by Dollaghan and Horner is that a positive likelihood ratio (LR+) of ≥ 10 is good. A LR+ of ≥ 3 is described as suggestive. On the other hand, a negative likelihood ratio (LR-) of ≤ 0.10 is informative while a LR- of ≤ 0.30 is merely suggestive (Dollaghan & Horner 2011). The positive likelihood ratios of studies determining the diagnostic accuracy of NWR tasks for assessing bilingual children varied from 2.74 (Guiberson & Rodriguez, 2013) to 12.50 (Boerma et al., 2015) indicating a range from suggestive to good. The results of the studies on the use of NWR tasks in the assessment of bilingual children are summarized in Table 3. Negative likelihood ratios ranged from .08 (Kohnert et al., 2006) to .62 (Gutiérrez-Clellen & Simon-Cereijido, 2010), indicating a good to poor accuracy level.

Most studies regarding the diagnostic accuracy of NWR tasks for identifying bilingual children with SLI reported better specificity than sensitivity, indicating that a high score on an NWR task could reliably rule out a diagnosis of SLI, while a low score could not reliably identify children with SLI. The LR+ values in the suggestive range are encouraging and indicate that these tasks warrant further investigation. The variability of results, however, indicates that the NWR tasks described are not yet sufficiently accurate for clinical use in isolation. LR- values are even more variable than the LR+ ratios. In general, the small number of studies using NWR tasks with bilingual children limits any conclusions that may be drawn regarding diagnostic utility.

Differences in Studies on Diagnostic Accuracy of NWR Tasks. Studies on the diagnostic accuracy of NWR tasks differ in key aspects, which may have impacted their findings and therefore caused the inconsistencies in diagnostic accuracy findings. Studies differed on the types of bilingual children examined (simultaneous or sequential), how the bilingual status of children was determined, how the language-impaired group was identified, and the age of participants. In addition, studies differed on which NWR task was used and how it was scored. Each of these factors is discussed in more depth below.

Participants. Child participants differed on their level of language exposure and language history. While many studies, especially those focusing on Spanish-English bilingual children in the US, included sequentially bilingual children, the children's language status was often not explicitly stated, making the comparison of studies difficult. In addition, details on the percentage of overall exposure or current use of English and the L1 were not reported. Researchers did not control for factors related to language exposure, a striking omission given research indicating the impact of language experience and vocabulary knowledge on NWR performance. One exception is Thordardottir and Brandeker's (2013) study, which explicitly stated that the participants were simultaneous learners of English and French and provided specific information on the relative exposure to each of the two languages, determining the total percentage of English exposure over each child's lifetime, including current usage.

Studies also differed on the age of participants, although almost all of the studies reviewed here focused on school-aged children (aged 5-11). Details on the age of the participants are listed in Table 4. One exception is Guiberson and Rodriguez (2013) who focused on preschool-aged children and found that NWR performance improved with age, replicating very similar accuracy rates found in studies conducted with school-aged children. As mentioned earlier, age is an important factor to consider because the relationship between language experience, vocabulary knowledge, and NWR performance changes with age. The age range for some of the studies is quite large, especially given the fact that the relationship between phonological memory and lexical knowledge with regard to NWR performance is even more complex in bilingual than monolingual children. In general, studies on bilingual children appear to target slightly older children than monolingual research, perhaps to allow for additional language exposure in the L2.

Tasks. Studies differed considerably on what NWR task or combination of tasks were used and how tasks were scored. Details regarding the types of tasks used are provided in Table 4. Some studies used an English-language task in combination with a task based in the child's L1 (Gutiérrez-Clellen & Simon-Cereijido, 2010; Windsor, Kohnert, Lobitz, & Pham, 2010) while others used only an English-based task (Kohnert et al., 2006; Paradis, Schneider, and Duncan, 2013). Most studies that used an English-based task used Dollaghan and Campbell's (1998) nonword repetition task. Tasks based in other languages, such as Spanish, French, and

Dutch, differed in their characteristics; only one study has examined the use of a commercially available, norm-referenced NWR task, the *Nonword Repetition* subtest of the *Comprehensive Test of Phonological Processing* (Paradis et al., 2013). Most of the NWR tasks included items that were 1-5 syllables in length and varied on phonological complexity. The tasks differed in key characteristics, such as wordlikeness, length, and articulatory complexity. These factors may influence a child's performance, thus impacting the diagnostic accuracy of the task. A meta-analysis such as that conducted by Estes, Evans, and Else-Quest (2007) for monolingual children has yet to be conducted for bilingual children. Therefore, no conclusions may be drawn regarding the relative diagnostic accuracy of the different NWR tasks utilized.

Scoring. Researchers employed both whole item scoring and percent consonants correct, often following the methods proposed by Dollaghan and Campbell (1998). This scoring method advocates that substitutions and omissions are scored as incorrect, whereas distortions and additions are scored as correct. Using percent consonant or percent phoneme correct scoring ensures the level of detail necessary for research purposes. However, most studies also explored the accuracy of whole item scoring, which would most likely be used in a clinical setting. In fact, Guiberson and Rodriguez (2013) determined that whole item scoring had better discriminating ability than percent consonant correct scoring. Very few studies note the types of substitutions that may be caused by L1 influence. One exception is Gutiérrez-Clellen and Simon-Cereijido (2010) who did not count Spanish-influenced articulation errors as incorrect. L1 influence is an important factor to consider when developing scoring procedures. Variations on the types of NWR tasks employed as well as differences in the children tested likely led to the variability of findings regarding NWR tasks' ability to correctly identify children with SLI.

Sampling. One of the limitations of most of the reviewed studies is that they utilized pre-determined groups of typically developing children and children with SLI. Dollaghan and Horner (2011) defined these studies as *two-gate designs*, in which the researcher recruits a group of children with SLI that has already been identified and another group of children who are typically developing. These studies are likely to introduce a *spectrum bias*, which would exist because the variability of a previously identified group with SLI would be lower than that occurring in the total population. In contrast, a *one-gate design* would simply draw a population sample and would separate the diagnostic groups after testing is completed. This would more accurately reflect the population because borderline cases would be included in the SLI group, leading to more precise estimates of diagnostic accuracy of NWR tasks. Details on study design of previous research regarding the diagnostic accuracy of NWR tasks are summarized in Table 5.

All of the reviewed studies, with the exception of Windsor et al. (2010), had a two-gate design. Some of these two-gate studies completed additional testing to confirm the SLP's prior assessment (Guiberson & Rodriguez, 2013; Gutiérrez-Clellen & Simon-Cereijido, 2010; Windsor et al., 2010), but others took no steps to verify the previous group determinations (Boerma et al., 2015; Girbau & Schwartz, 2008; Kohnert et al., 2006; Linden, 2009; Paradis et al., 2013; Thordardottir & Brandeker, 2013). In Windsor et al.'s (2010) one-gate design study, children were recruited from a school setting and thus included those with and without SLI, who were identified though their participation in speech-language services in the school. While the status of the children with SLI was verified, no additional testing was conducted to rule out language impairment in the typically developing group. Because not all children were given the same assessments, the recruitment methods do not rule out a possible bias in determining the children's language status. Using a one-gate design is often prohibitive due to the large number of participants required. In fact, Windsor et al. (2010) included an impressive 187 children in their study. It should be noted that the likelihood ratios found by Windsor et al. (2010) were similar to those found by studies with a two-gate design.

A related methodological concern is that only one study, Thordardottir and Brandeker (2013), reported the use of independent testing, using different examiners for the two languages tested. In addition, none of the reviewed studies indicated the use of blinded testing, for which the examiner was not aware of the child's language status. Independent and blinded assessment of a child's language status as typically developing or SLI is the key to conducting a study on the accuracy of using NWR tasks as a diagnostic tool. It is only by comparing the rates of identification that the accuracy of the task may be checked.

Conclusion Regarding Previous Literature

Monolingual children's performance on NWR tasks accurately identifies those with SLI. However, due to factors related to learning two languages, such as age of first exposure, amount of exposure, or phonological characteristics of the child's languages, a bilingual child's performance on a NWR task varies considerably, thus complicating their use as a diagnostic tool. While a thorough assessment of both of a child's languages is recommended, it is often not practically possible for monolingual SLPs. Therefore, it is necessary to find another method of assessment for this population. One possibility is utilizing a processing-based measure, such as a NWR task. While these tasks are quite accurate when used with monolingual children, their accuracy in identifying bilingual children with SLI varies greatly based on the task design and population studied.

A review of the literature revealed that very few studies specifically addressed the diagnostic accuracy of NWR tasks for use with bilingual children. Those that did address diagnostic accuracy varied on several aspects of task and study design. All of the studies except one represented a two-gate design, making them susceptible to spectrum bias. Thus, research studies should include a one-gate design. With a one-gate design, it would be possible to utilize the statistical analysis method of quantile regression in order to determine the nature of the relationship between variables. Quantile regression utilizes all data points, decreasing the loss of power that occurs when dividing participants into diagnostic groups. In addition, this statistical method provides a more nuanced interpretation of the relationship between variables. For example, it is possible that an ordinary least squares (OLS) regression would not show a relationship between variables, but the quantile regression may indicate a change in the strength of relationship across quantiles of the outcome variable. Future research should explore the use of this statistical method. In addition to a problematic study design, prior studies were missing key information regarding the language background of participants, such as whether participants underwent identical testing procedures and whether researchers were blinded to the diagnostic status of participants. Most of the studies used parent and teacher report of concerns as one way to reference a child's language status, but many did not verify language status with independent tests. Thus, future studies should include independent testing, blinded testing, and consistency in the tests given to participants (e.g., all participants are tested in the same manner, regardless of diagnostic status). Finally, studies did not provide sufficient background information on participants regarding language exposure and experience, which could influence NWR task performance. Studies should provide ample background information regarding the participants' language experience. Ideally, a range of experience should be represented in the participants. Research should determine the role of language experience in task performance, in an effort to determine which task is least influenced by a child's language experience, thus maximizing the difference between typically developing children and those with SLI. In conclusion, no specific nonword repetition task has stood out as being the best diagnostic marker for bilingual children although the "quasi-universal" task utilized by Boerma et al. (2015) showed the most potential. Tasks were accurate in ruling out a diagnosis of SLI for children whose NWR performance was high. However, when task performance was low, tasks could not rule out a diagnosis of SLI.

Purpose of Current Study

The purpose of the current study was to examine how children's performance on three nonword repetition tasks (NWR) that differ in theoretical design relates to SES, language experience, and language ability. In addition, the study explored the predictive power of the three NWR tasks in identifying children with low language ability. The children's performance on each of the NWR tasks, as defined by percent phonemes correct, was examined based on their SES, measured as parental years of education, and percentage of English exposure, as determined by parent questionnaire. In addition, performance on the three tasks was examined related to the children's ability in Spanish and English, based on their performance on the *Bilingual English Spanish Assessment* (BESA). Finally, this study explored the predictive power of each of the NWR tasks in identifying those children who scored at the 16^{th}, 25^{th}, or 50^{th} quantile on the BESA to address whether an OLS regression accurately represents the relationship between NWR performance and the language measure.

Research Questions

The following questions were addressed in this study:

1a. How do children perform on three differently designed NWR tasks and what is the pattern of their performance?

1b. Does SES, measured as parental years of education, impact the nonword repetition scores on three differently designed NWR tasks?

1c. Does age of first exposure (years from 0-6) impact NWR performance?

2. When controlling for SES/age of first exposure (if necessary), what is the predictive power of percentage of current English use (input and output, averaged) and performance on the three NWR tasks?

3. For each NWR task, what is the task's predictive ability in identifying children with language scores at the 16th, 25th, and 50th quantile of the BESA Index score?

It was predicted that the three tasks would be significantly correlated, as they all should measure the underlying construct of phonological memory. No significant differences between tasks at any syllable level were predicted. SES was not predicted to impact scores on the three NWR tasks. Similarly, age of first exposure was not predicted to relate to children's task performance. For research question 2, the researcher predicted that the children's language status, expressed as a percentage of time that English is heard and spoken, would be statistically significantly related to their performance on the norm-referenced task based in English (CTOPP-2). The researcher predicted that the relationship between the other two tasks, the "Quasi-universal" (QU) and the "language neutral" (LN) task, would not be statistically significantly related to the children's language status.

For research question 3, it was predicted that the strength of the relationship between the CTOPP and the BESA would not be statistically significantly different at the 16th, 25th, or 50th quantile. In other words, the relation between the CTOPP and the BESA score would not be conditional on the percentiles of the BESA, indicating that the relationship between the CTOPP and the BESA would be consistent across quantiles. This was predicted because it was hypothesized that the CTOPP performance is related to language experience and other outside factors rather than language knowledge. For the QU and the LN tasks, it was predicted that the relationship between the variables at the16th percentile would be significantly stronger than at higher quantiles, indicating that the relationship between the NWR task and BESA would be stronger for lower quantiles than higher quantiles. This would indicate that these tasks reveal a stronger relationship between the variables at lower NWR task performance and thus have potential as diagnostic measures.

Chapter 3: Method

Participant Recruitment

As an exploration into the potential clinical utility of three different nonword repetition (NWR) tasks, the current study addressed several limitations of previous research regarding the diagnostic accuracy of NWR tasks. Most previous studies relied on a two-gate design, drawing from a group of children previously identified as having Specific Language Impairment (SLI). Such a design may lead to a *spectrum bias,* which results from the fact that the group with SLI may not exhibit the same variability as a population sample, leading to the overestimation of the diagnostic accuracy of the assessment protocol (Dollaghan & Horner, 2011). The current project addressed this limitation by drawing a sample of children with researchers blinded to each child's language status during assessment sessions. Instead of examining children based on their language status (language impaired or typically developing), this study examined the ability of the NWR tasks to predict the children's performance on a bilingual assessment that has been normed on bilingual children: the *Bilingual English Spanish Assessment* (BESA). All of the children, regardless of prior language status, were evaluated using this measure as well as other measures of phonological ability, hearing, and nonverbal intelligence. Not using previously-identified children is especially important given the fact that an estimated 70% of children with SLI are not yet identified at the age of five (Tomblin, Smith, and Zhang, 1997). Children were recruited from elementary schools and community locations in Lawrence, KS, Kansas City, KS, Kansas City, MO, Topeka, KS and surrounding areas. Children were between the ages of 5;0 and 6;11 at the time of testing and were bilingual in Spanish and English. This age range is consistent with the literature on monolingual children with SLI. In addition, the first elementary years are crucial in the early identification of children with SLI.

Participants

Participant characteristics are outlined in Tables 6 and 7. Fifty-nine participants were recruited, and forty-seven participants were included in the final analyses. Attrition of participants was due to several factors, including lack of follow-up by parents after initially enrolling their child (6 participants), not meeting inclusionary criteria (5 participants) and loss due to an error in task administration (1 participant). Children, (21 female, 26 male), were drawn from preschool, kindergarten and first-grade classrooms in the Lawrence/Topeka/Kansas City areas as well as from community locations such as public libraries or church communities. The majority of participants (60%) were enrolled in kindergarten classrooms. While research on bilingual children often includes a much broader age range, this age range is supported by the literature on monolingual children with SLI. An elementary age range is appropriate for a focus on early identification of children with SLI, which is crucial in preventing the long-term educational impact of under-diagnosis (Kohnert, 2013). Children were simultaneous or sequential language learners. Because the BESA is normed on bilingual children with various ages of acquisition, a child's age of first exposure should not impact their BESA performance. All participants were screened for inclusionary criteria: no prior diagnosis of any neurological condition (e.g., ADHD, ASD, seizure disorder), passing a hearing screening at 25dB at 1,000, 2,000, and 4,000 Hz in both ears, and having between 20-80% of exposure (receptive and expressive, averaged) to English and Spanish in daily activities, as assessed per parent report. In addition, all children had a nonverbal IQ score within typical limits (>16th %ile). These exclusionary criteria are consistent with the literature and definitions of SLI.

A parental consent form providing a Spanish and English summary of the study purpose and tasks involved was given to all of the parents. Once testing was completed, parents were provided with a report, in Spanish or English, based on parent preference, summarizing their child's performance on all of the assessments conducted. This report included a form to request that additional copies be sent to the child's school or physician to be used in further educational decisions and planning.

Shown in Table 6 are the participant characteristics. The average age of the participants was 5;11, with a range from 5;0 to 6;11 (SD= .58). The average percentage of English heard and spoken by the children was 49% percent (SD= 18), with a range from 20 to 80. The parent education level, measuring a child's SES, was an average of 12 years (SD= 4), with a range from 6 to 23 years. The average Nonverbal Index Score from the Reynolds Intellectual Assessment Scales was 106 (SD= 14), with a range from 86 to 140. Shown in Table 7, the average age of first exposure to English was 2 years (SD= 2), with a range from 0 to 5 years of age. Most of the participants reported that their ethnicity was Hispanic/Latino (41 out of 43 participants, with four not reported). In addition, most of the participants indicated that they spoke a Mexican dialect of Spanish (23 Mexican, 21 not reported). Other dialects reported were Castilian (2), Argentine (1), Bolivian (1), Salvadorian (1), Cuban (1), and Ecuadorian (1).

Children were recruited to represent the continuum of relative amount of exposure to English, and the entire group was included in the analysis on the effect of exposure on performance. Children's exposure to English was calculated through the use of a parent questionnaire, an adapted version of the Bilingual Input-Output Survey (BIOS), which is part of the protocol of the Bilingual English Spanish Assessment (BESA). This questionnaire surveys language use in the home and in school or daycare for an average weekday and weekend day. Questions on exposure ask the parents to list the child's communication partners, which language they speak, and how much they interact with the child. In addition, for each year of life, the parents listed daycare settings attended and language(s) spoken. Other settings of significant exposure are reported as well (such as regular weekends with grandparents). From these data, a single number was computed representing the percentage of the child's waking hours spent in exposure to English (receptive and expressive)—the remaining waking hours being spent in exposure to Spanish as no other languages were part of these children's environment on a regular basis. The questionnaires used, as well as the scoring method, are attached in Appendix A.

Procedures of Data Collection and Training

Data were collected by the primary researcher, a bilingual Spanish-English Speech-Language Pathologist, and trained graduate and undergraduate students in Speech-Language-Hearing. The principal investigator provided direct training to all data collectors. Training involved familiarization with task protocols, practice of task administration, and practice in scoring the assessments. The researcher trained students by having them first observe at least two testing sessions, asking them to run two testing sessions while supervised, and then independently assessing children. All assessment sessions were reviewed using video or audio recording to ensure consistency of testing procedures. All testing sessions were audio and video recorded for analysis and reliability scoring. All researchers involved in data collection were blinded to the participants' language status. This latter point is crucial because many prior studies did not include blinded researchers, a point that may have introduced testing bias. In addition, a different researcher assessed each language. In other words, if a researcher conducted the English language testing, the same researcher would not conduct the Spanish session. This is important because having a unique researcher for each language will decrease the amount of codeswitching in which the child engages.

Test protocol. After parental consent was obtained, all measures were administered to all participants. The following assessment measures were given across three sessions, each lasting approximately 30-50 minutes: Reynolds Intellectual Assessment Scales (RIAS), hearing screening, Bilingual English Spanish Assessment (BESA)-English, Bilingual English Spanish Assessment (BESA)-Spanish, Comprehensive Test of Phonological Processing, 2nd Edition (CTOPP-2) Nonword Repetition subtest, "Quasi-universal" (QU) NWR, and "Language Neutral" (LN) NWR. All of the assessment measures were administered to the children in a quiet room in their school, a community space, or in the home.

Table 8 outlines details regarding the structure of the assessment sessions and how tasks were randomized across sessions. Each session began with the researcher eliciting a verbal assent from the child. The assent was given in the testing language (Spanish or English). The order of sessions was counterbalanced across participants. However, the RIAS and the hearing screening were always administered in the first session, as these were the exclusionary criteria for study participation. The NWR tasks were randomized across participants but the CTOPP was always presented during an English-language session (i.e., it was presented with either the RIAS/hearing screening or the BESA-English). Because the CTOPP is based only in English, presenting it within a Spanish language context (i.e., a Spanish language assessment session) could have influenced test results. Children were offered one break during each session and were encouraged through positive verbal reinforcement as well as tangible rewards (e.g., stickers and small prizes).

All NWR tasks were administered with at least 16 items presented to the child. The administration of the CTOPP differed slightly from the procedure outlined in the standardized manual. The CTOPP nonword repetition task was administered until a child missed three items in a row (the ceiling set by the standardized test). However, if a child hit a ceiling score before completing at least all items that are 1-5 syllables in length, the examiner continued until all items from 1-5 syllables in length were administered. This difference is justified because it allowed for both the standardized scoring of the measure as well as a more direct comparison of the NWR tasks. Because the CTOPP items are not presented from shortest to longest, more than 16 items were included in the presentation. All children completed at least 19 items on this test. The "quasi-universal" and the "language neutral" tasks contained 16 items from 2-5 syllables in length.

The NWR tasks were administered as part of the larger assessment session. In an "English" session, the researcher spoke only English with the child; in the "Spanish" session, the researcher spoke only Spanish with the child. NWR repetition items were presented in semi-random order for the QU and LN task using the Direct RT software to randomize presentation of items for each syllable length (Jarvis, 2014). Items for the CTOPP were presented in the standardized order. Items were presented over an external speaker that was adjusted to the participant's loudness preference. The listening volume was tested before each testing session, with the researcher asking the child whether the presentation was loud enough. Nonwords were presented only once according to the conventional administration procedures outlined in the CTOPP, and the children repeated the item immediately after hearing it. Digital recordings of children's repetitions of the nonwords were made during each session.

Nonword Repetition Tasks Utilized

The Nonword Repetition subtest of the Comprehensive Test of Phonological Processing—2nd Edition (CTOPP-2) (Wagner, Torgesen, & Rashotte, 1999). This is a measure of phonological short-term memory. It requires the child to listen to and immediately repeat nonsense words played on a CD. These items increase in length and phonological complexity. Children are scored on how many items they accurately produce until they reach a ceiling of errors, with omissions and substitutions counting as errors. The standard score of the subtest has a mean of 10 and a SD of 3. For the purposes of the study, children were required to complete all items from 1-5 syllables in length, regardless of ceiling scores. Table 9 lists the length and characteristics of test items. Subsequent transcription and scoring of the test for the purposes of this research study included a score using percent phoneme correct.

Quasi Universal Nonword Repetition Task. The "quasi-universal" (QU) task, developed by Chiat (2015) as part of the COST Action IS0804 team in Europe, was developed with the following characteristics in mind: item length, prosodic structure, segmental complexity, and phonotactic probability. These constructs are known to affect nonword repetition performance in English and affect children with and without SLI according to the properties of real words that they have experienced. Table 10 outlines the item characteristics for this task. While it is impossible to develop a truly "universal" task, choices were made according to which characteristics would be maximally applicable across the world's languages. This test contains items from 2-5 syllables in length (4 items at each length). Consonants were limited to those most commonly found in the world's languages: /p, b, t, d, k, g, s, z, l, m, n/ and vowels were limited to the tense vowels /a, i, u/. The QU task includes one sound not found in the Spanish inventory, /z/. The task items included syllable structures that are valid in both English and Spanish. Items were developed to follow a CVCV structure, making them compatible with the structure of most of the world's languages. Items were produced using the phonetic realizations of the language that is dominant in the child's environment (English, in this case). The task was recorded by a native speaker of English, who produced each item using the phonotactic realizations of English, with equal stress placed on each syllable. The task was recorded in a manner that is consistent with other nonword repetition tasks.

Language Neutral Nonword Repetition Task. This task was developed using the phonemic inventories of English and Spanish. All sounds unique to one language were eliminated. Next, early-developing phonemes common to both languages were combined to produce CV pairs. This included the phonemes / m, n, w, p, b, t, d, k, g, f, h/. A list of 2-syllable CV words using these phonemes was created, of which real words or dialect words of Spanish were eliminated with the help of two Spanish-speaking informants. All real English words were also eliminated with the help of two informants. The remaining items were constructed using the original two-syllable items. For each set, the informants rated the items' "wordlikeness," described on a scale from 1 (not at all wordlike), 2 (neutral), to 3 (wordlike). Only items designated as 1 (not at all wordlike) were included in the stimuli. The researcher randomly selected a total of 4 stimuli per syllable length (2-5) to be recorded. Stimuli were recorded by a native speaker of English, who produced each item naturally, placing equal stress on each of the syllables. See Table 11 for a list of items included in the task.

For the LN and QU task, the stimuli were manipulated in the PRAAT software to standardize them for intensity and length (Boersma, 2001). A blank period was added to each nonword recording to allow time for the child to respond, making the task consistent with the CTOPP-2 recordings. Items were analyzed to determine how long item time was across tasks. The average duration of a non-word for all tasks at each length was: two-syllables (1.03 seconds), three-syllable (1.36 seconds), four-syllable (1.45 seconds), and five-syllable (1.94 seconds).

Task Scoring

Scoring of the CTOPP was completed (1) according to the standardized procedure and (2) according to the guidelines developed by Dollaghan and Campbell (1998). The standardized procedure indicates that an item is correct when all of the phonemes of the item are produced correctly. Any missing sounds, additional sounds, or sounds out of order result in a score of zero for that item. Dollaghan and Campbell's (1998) scoring method includes a *whole item score* as well as a *percent phoneme correct* scoring. Any substitutions, omissions, or changes in the order of phonemes were scored as incorrect. For example, if the target was /zibu/ and the child produced /tibu/, this would be scored as incorrect. However, distortions are scored as correct, as were "accented speech" that were also present in a child's spontaneous speech sample and on the standardized articulation/phonology tests. Additions were ignored. This ensured that a child was not punished for producing "accented" productions of the nonwords. The QU and LN tasks were also scored in this manner. Scoring the tasks in a *whole item* manner as well as a *percent phoneme correct* enabled the researchers to draw conclusions regarding which scoring method would have the most clinical relevance. Research on scoring methods has shown differing results, with several researchers indicating that a *whole item* score was more accurate or had similar accuracy in identifying children with SLI (Guiberson & Rodríguez, 2013).

Reliability

For the purpose of reliability checking, all of the children's NWR tasks were transcribed phonetically by a second trained judge to obtain a measure of inter-rater reliability. Reliability for each task transcription was as follows: for the CTOPP (96%), the QU task (97%), and the LN (96%). Any discrepancies in task transcription and scoring were addressed by a third transcription, which was used to determine the final scoring of each item. A second judge calculated all scores on the BESA and the *Reynolds Intellectual Assessment Scales* (RIAS) in order to maintain the accuracy of test scoring. In addition, 20% of all tests (BESA English, Spanish, and RIAS) were scored again to determine reliability of initial interpretation of a child's response. For the NWR tasks, another researcher scored 20% of the calculations of percentage of phonemes correct in order to determine the reliability of the original scoring. For the scoring of percentage English for the participants, the inter-rater reliability was 81%. While this reliability is somewhat low, the discrepancies between scores were all less than 5%. For the percentage of phonemes correct for the LN task, QU task, and CTOPP task, the reliability was 99%, 97%, and 97%, respectively. For the BESA English test, reliability was 100%; for the BESA Spanish test, reliability was 90%. For data entry, the reliability was 100%.

Bilingual Language Test Performance

Children's performance on the BESA varied considerably, with some children scoring better in Spanish than in English. However, most participants were dominant in English, a pattern that is expected because all of the children were tested in an English dominant environment (i.e., central region of the United States).

Descriptive data for all of the assessment measures are presented in Table 12. All reported scores are standard scores, with a mean of 10 (SD= 3) for the subtest scores (phonology, cloze, sentence repetition, receptive, and expressive vocabulary) and 100 (SD= 15) for the corresponding combined scores (morphology and semantics). The morphology composite score comprises the combined standard scores for the cloze and sentence repetition subtests. The semantic composite score is composed of the receptive and expressive vocabulary scores. These combined scores are *italicized* in Table 12. The overall index score combines the highest score on the morphology and the semantic section, regardless of language. Thus, this score represents children's overall language ability, and a low index score indicates language impairment (Paradis, 2015). In other words, although many bilingual children have mixed language profiles, this method of scoring accurately reflects a child's language ability when considering both of the children's languages. Thus, children with typical language skills will score within typical limits in their stronger language, while children with language impairment will score below typical limits on test composites in both languages. For the purpose of this study, the BESA Index score was also calculated using only the English subtest scores ("English Index") and using only the Spanish scores ("Spanish Index") in order to estimate a child's language ability in each language separately.

To examine which language had the higher performance, the subtests and combined scores were compared using paired samples t-tests. Comparing the Spanish and the English performance on each of the subtests revealed a significantly higher mean on the English subtests, which is not surprising given the fact that data were collected in the United States and that most of the children attended English-only schools. A paired-samples t-test revealed that there were significant differences between the children's performance on the English and Spanish morphology sections ($t(46) = -4.97$, $p < .001$) and semantics combined scores ($t(46) = -2.29$, $p = .03$) but not the phonology subtest ($t(46) = -1.13$, $p = .27$). This is logical, given the fact that a child's phonological ability is a skill that would be less affected by language exposure and should thus be highly consistent across languages. On the other hand, semantic and morphological skills would be impacted by the amount of exposure to each language. This relationship between languages was explored further, as outlined below. The cloze and sentence repetition subtests correspond to the children's morphological ability in Spanish. To examine differences across languages on the individual subtests, additional paired samples t-tests were conducted. The individual subtests revealed a significant difference on the cloze test (t (46) = -5.10, P< .001), sentence repetition (t (46) = -3.98, P< .001), and expressive vocabulary (t (46) = -2.10, *p* = .04), all favoring scores on the English tests. It is predictable that the children's receptive language does not differ statistically across languages, because receptive language skills tend to be more equal in bilingual children, whereas the expressive skills reveal the child's dominant language (Bedore et al., 2012). Therefore, although the children in this sample did not differ in their receptive skills in Spanish and English, their English language dominance was revealed in their expressive language ability.

This pattern of language dominance is supported by further analyses. The BESA Index (using the strongest performance on the morphology and semantic subtest, regardless of language) and the BESA Index based on only the English scores were also significantly different (t (46) = 4.12, p <.001), with the BESA English score being higher. In fact, for all of the subtests assessed, the English language scores were higher than the Spanish scores, although not always significantly so, indicating that the children sampled for this study generally have higher language skills in English across language domains.

Chapter 4: Results

This study examined bilingual children's performance on three nonword repetition (NWR) tasks. It explored children's performance on these tasks related to children's socio-economic status (SES), age of first exposure, and how task performance relates to language ability. The study answered the question of how a child's language experience (i.e., relative exposure to each language) influences performance on NWR tasks and how well tasks correlated with children's performance on a standardized assessment in each of their languages. The three tasks differed in the extent to which their nonwords contained sound patterns found in the test language, English or Spanish. These research questions determined which task has the most potential to be used as a diagnostic tool for language impairment in bilingual children. A description of the main research questions and results follows. Analyses were conducted with a complete data set for 47 participants, with no missing data points. Before addressing the main research questions regarding how tasks are influenced by language experience and how they related to the children's language ability, the participants' performance are explored descriptively.

Research Question 1

Research Question 1a. How do children perform on three differently designed NWR tasks and what is the pattern of their performance?

Before addressing the other two research questions regarding each task's potential for diagnosing bilingual children with language impairment, the pattern of results for the three NWR tasks was explored. Tasks were compared at each syllable level to determine whether statistically significant differences occur. It was predicted that task performance would decrease with increasing item length but that there would be no statistically significant differences between tasks. All scores presented below are in percent phonemes correct. Figure 1 shows the average scores on the NWR tasks, which decrease with increasing syllable lengths. This is expected because longer items place a larger burden on the child's phonological memory (Dollaghan & Campbell, 1998). Comparisons between task performance were conducted using one-way ANOVA. This revealed statistically significant differences between tasks at all syllable levels. Results showed significant differences at the two-syllable level ($F[2, 45] = 38.79$, $p < .001$), the three-syllable level ($F[2, 45] = 46.08$, $p < .001$, the four-syllable length ($F[2, 45] = 34.69$, $p < .001$), and the five-syllable length ($F[2, 45] = 22.25$, $p < .001$). For all of these comparisons, the QU task performance was higher (i.e., children were more accurate on this task) than either of the other tasks at all syllable lengths. In contrast, the children's performance on the CTOPP was generally worse than the other two tasks at all syllable levels. Follow-up analyses revealed that the QU and the CTOPP tasks differed significantly at all syllable lengths. The CTOPP and the LN task differed significantly at the two- and four-syllable level, and the QU and the LN task differed significantly at the three-, four-, and five-syllable level. Table 13 provides details on follow-up testing for each task at each syllable level. Further analyses, discussed below, determined whether this task differed from the others with respect to its relationship with language experience or ability.

Research Question 1b.Does SES impact the nonword repetition scores on three differently designed NWR tasks?

This question examined whether the tasks were biased toward children's SES standing. It was predicted that SES would impact none of the tasks. Prior research has shown that NWR tasks are not influenced by a child's SES, a key point that distinguishes these types of tasks from knowledge-based measures, such as vocabulary (Hoff, 2003). In order to answer this research questions, three regressions were conducted, with SES as the independent variable and the three NWR tasks as the dependent variable. SES was calculated using the parents' highest level of formal schooling, expressed in years of education completed. This quantification of SES is commonly utilized in the literature on language learning in monolingual and bilingual children. If only one parent's information was given, this was used; if both parents' level of education was given, the higher of the two was used. Years of education were coded numerically on a scale from 6 to 23, based on parent report. In other words, a high school degree (or equivalent) was coded as 12 years, and a 4-year college degree was coded as 16 years. The largest percentage of parents (26%) indicated completing a high school degree. As predicted, no significant relationship was found between SES and any of the tasks. For the LN task, ($F[10, 36] = .80$, $p = .63$), for the QU task, ($F[10, 36] = .42$, $p = .93$), and for the CTOPP, ($F[10, 36] = .61$, $p = .79$). Therefore, SES was not controlled for in any further analyses.

Research Question 1c. Does age of first exposure (years from 0-6) impact NWR performance?

Age of first exposure (AFE) was not predicted to impact task performance on any of the tasks. Three regressions were run with AFE as the covariate and the three tasks as the dependent variables. Testing revealed a significant relationship between AFE and the performance on the LN task ($F[5, 41]= 3.09$, $p = .02$) but not the other two tasks: for the QU task ($F[5, 41]= 1.82$, $p = .13$) and for the CTOPP task ($F[5, 41])= 2.00$, $p = .10$). Because this result was unexpected, further analyses concerning the NWR tasks and the BESA test scores controlled for AFE.

Research Question 2

When controlling for age of first exposure, what is the predictive power of percentage of current English use (input and output, averaged) and performance on the three NWR tasks?

It was predicted that the current usage of English would impact the CTOPP, which is based on English phonology, but not the LN or QU task. Three linear regressions were run. In a simultaneous regression, the percentage of phonemes correct was regressed on age of first exposure and percentage of English. For each task, results for the overall regression (with both independent variables) and then the individual regression (for each variable separately) are presented below. The regression for percentage English and the tasks represents the impact of percentage English when controlling for AFE.

For the LN task, the prediction that the task would not be impacted by language experience was not supported by the analyses. Testing revealed that the two dependent variables together (age of first exposure and percentage English) explain roughly 24% of the variance in students' performance, a relationship that is statistically significant ($F[2, 44] = 6.92$, $p = .002$). However, the individual variables (AFE and percentage English) were not significantly associated with the performance on the LN task. For AFE, testing revealed ($\beta = -.18$, $t(46) = 6.53$, $p = .33$). Similarly, for percentage English, ($\beta = .35$, $t(46) = 1.89$, $p = .07$). The effect of percentage of English is approaching significance and has a large effect size, suggesting that statistical power may not have been sufficient to detect the effect. Future research with greater power is needed to determine whether the percentage English significantly influences performance on the LN, but a preliminary conclusion is that the task may not be as linguistically neutral as planned. It is also possible that another underlying factor may influence performance on this task, a possibility that will be explored further below.

For the Quasi Universal task, the hypothesis that the task would not be impacted by language experience was supported by the data. The analyses of the QU revealed ($F[2, 44] = 3.18$, $p = .051$), indicating that the combined effect of AFE and percentage English is approaching significance. However, the individual variables are not statistically related to task performance, indicating that some other factor may be influencing children's performance on this task. For AFE, testing revealed no significant relationship (β= -.19, *t* (46) = -.95, *p* = .35). Similarly, for percentage English, no significant relationship was found (β= *.20, t(46)= 1.02, p= .32)*. Thus, unlike the LN task, the QU task performance is not significantly influenced by language experience, suggesting that this task may be a better assessment tool for phonological memory in bilingual children.

Finally, for the CTOPP, a picture similar to that of the Language Neutral task emerges, with an overall statistically significant effect (F[2,44] = .19, *p* = .01). For AFE, testing revealed no significant relationship (β= -.10, *t* (46) = -.50, *p* =.62). For percentage English, the relationship between percentage English and the task performance approached significance (β= *.346, t(46)= 1.95, p= .057).* Similar to the LN task, the effect size is large, suggesting that the analysis did not have sufficient statistical power to reach significance. These results indicate that performance on the CTOPP task is nearing statistical significance in its relation to percentage of time listening to and speaking English, when controlling for age of first exposure to English. Future research should include a larger number of participants in order to increase the statistical power of the analysis. The current analyses suggest that the CTOPP performance is significantly impacted by current language use and thus functions similarly to the LN task.

In summary, it was expected that the percentage of English, after controlling for age of first exposure, would be significantly related to the performance on the CTOPP but not on the other two tasks. This hypothesis was not supported by the data, which showed that the percentage of English was approaching significance in its impact on the CTOPP and the LN task performance but not for the QU task, when controlling for AFE. While the results are approaching statistical significance for the LN or the CTOPP, the QU task revealed no significant impact of percentage of time spent speaking and listening to English. Thus, this task may be less influenced by current language use and may therefore be more accurate in reflecting a bilingual child's phonological memory.

Further Analyses in Answering Research Question 2

Because the results for question two were unexpected, and some underlying factors of the LN and the CTOPP tasks were likely impacting performance, further analyses were conducted. One possibility that must be ruled out is that the task performance on all tasks is driven by some other factor, such as the child's articulation of speech sounds. Therefore, a regression was conducted, with the phonology subtest of the BESA as the independent variable, and each NWR task as the dependent variable. Results indicated that the phonology subtests on the BESA (in Spanish and English) and performance on some of the NWR tasks were significantly related, indicating that children's phonological skills may impact NWR task performance. For the Spanish phonology and the LN task, the relationship was approaching significance $F[33, 13] = 2.15$, $p = .07$. However, for the Spanish phonology, the relationship was not significant with the CTOPP ($F[29, 17] = 2.24$, $p = .16$) or for the QU task ($F[28, 18] = 1.51$ $p = .18$. For the English phonology subtest, the relationship between phonology and the CTOPP performance was significant ($F[29, 17] = 2.24$, $p = .04$). For the LN and the QU tasks, this relationship was not significant: for the LN task ($F[33, 13] = 1.66$ $p = .17$); for the QU task ($F[28, 18] = 1.64$, $p = .14$). In other words, the participants' English phonology ability may have impacted children's performance on the CTOPP task, whereas their Spanish phonology may have impacted performance on the LN task.

Since a significant relationship was found between the phonology subtests and the nonword repetition task performance, a differential scoring method was used to control for the impact of articulation skills on NWR task performance. In this developmental scoring method, the nonword repetition task accuracy was adjusted to account for children's articulation errors. The researcher determined all of the speech sounds that were produced in error during the BESA testing and took note of what kinds of errors were found (e.g., deletion of sounds in certain positions, substitutions of sounds, etc.) for each speech sound. Any of the speech sounds that were produced with the same phonological process (e.g., deletion or substitution), were then counted as correct in the developmental scoring of the nonword repetition tasks. To further analyze the pattern of performance with the developmental scoring method, three regressions were conducted, with percentage of English and age of first exposure as the predictors and the developmental scoring of each NWR task as the dependent variable. For the Language Neutral Task, the impact of AFE and percentage English combined was statistically significant (F[2, 44] = 6.74, *p*= .003). Similar to the findings with the original scoring, the effect of percentage English when controlling for AFE approached significance (*β*= = .34, *t* (46) = 1.84, *p* = .07). For the QU task, a similar picture emerged as the regular scoring F[2,44] = 260, *p* = .09, indicating that the overall impact of AFE and percentage English was not statistically significant for the QU task. Finally, for the CTOPP, testing indicated that the overall effect of AFE and percentage English on the CTOPP was statistically significant, F[2, 44] = 5.13, *p* = .01. Examining the impact of percentage English when controlling for AFE approached significance (*β*= .34, *t* (46) = 1.79, *p* = .08).

The picture that emerged from these analyses was very similar to that of the initial task scoring, which did not control for articulation errors. For the developmental LN task, the overall impact of percentage English and age of first exposure was significant, but neither variable individually had a significant impact on task performance, although the impact of percentage English was approaching significance. For the QU task, neither the combined impact of AFE or percentage English, nor their individual contributions were significant. For the CTOPP task, the developmental scoring mirrored that of the original scoring, which did not control for articulation errors. The combined impact of AFE and percentage English was significant, but neither variable made a significant contribution to task performance, with the impact of percentage English approaching significance. Using the developmental scoring did not change the overall pattern of results. Accounting for articulation errors did not remove the influence of language experience on test performance.

Research Question 3

For each NWR task, what is the task's predictive ability in identifying children with language scores at the 16th, 25th, and 50th percentile on the BESA?

In order to determine which of the three NWR tasks has potential as a diagnostic tool for diagnosing language impairment in bilingual children, a quantile regression was estimated for each of the NWR tasks for the quantiles .16, .25, and .50. These quantiles were chosen because they represent commonly used clinical markers for language impairment. A quantile regression determined whether the relationship between each NWR task and the BESA is consistent across BESA scores or whether the NWR task performance predicts differently at different quantiles. It was predicted that the relationship between the QU and LN tasks and the BESA would be stronger for lower quantile on the BESA. In other words, a low performance on the NWR task would be more strongly related to a low score on the BESA than an average score on the NWR task. A summary of quantile regression is presented in the Appendix. Before examining the quantile regression, the comparison with the Ordinary Least Squares (OLS) was established.

This allowed a comparison between the two methods of analysis. When conducting the OLS regression, three dependent variables were examined in relation to NWR performance: the BESA English Index (corresponding to the participants' language ability in English), the Spanish BESA Index (corresponding to the participants' language ability in Spanish), and the BESA Index (corresponding to the participants' strongest scores in either language). It was predicted that all three tasks would be predictive of the BESA Index score and the BESA English Index, but that only the LN and the QU tasks would be predictive of the BESA Spanish Index. The CTOPP would not be predictive of the BESA Spanish Index because this NWR task is based solely on English speech sounds. These predictions were supported by the regression results.

All of the NWR tasks were significantly related to the BESA English score, even when controlling for age of first exposure. For the LN task, the NWR task was statistically significantly related to the BESA English Index (F[2, 44] = 35.87, p <.001). Based on the standardized coefficient, a 1 SD increase in the LN task performance would lead to a .6 SD increase on the BESA, a change that is statistically significant and corresponds to a large effect size *(β= .60, (t(46)= 6.88, p< .001)*. For the QU task, the overall impact of AFE and percentage English on the QU task is statistically significant (F[2, 44] = 18.33, p< .001). Examining the impact of the QU task on the BESA English Index, when controlling for AFE revealed that increase in 1 SD increase on the QU task would lead to a .52 SD increase in the BESA score, a difference that is statistically significant and corresponds to a large effect *(β= .52, t(46) = 4.43, p< .001)*. Finally, for the CTOPP task, the overall relationship is statistically significant (F[2, 44] = 36.77, p< .001). For the CTOPP task performance, when controlling for AFE, the relationship between the CTOPP task performance and the English BESA Index is statistically significant *(β= .69, t(46) = 6.98, p< .001)*. Based on the standardized coefficient, a 1 SD increase in the CTOPP score would lead to an increase of .69 SD increase on the BESA, an effect size that can be characterized as large. Therefore, when controlling for AFE, all of the NWR tasks are significantly related to the English Index score on the BESA.

For the BESA Spanish Index, a somewhat different pattern was found, one that follows the prediction that the CTOPP would not be predictive of the Spanish language performance. Using a simple regression, each NWR task was regressed on the Spanish Index BESA score. When using the BESA Spanish Index score, only two of the NWR tasks were significantly related to the BESA score, when controlling for age of first exposure. For the LN task, the NWR task was statistically significantly related to the BESA Spanish Index ($F[2, 44] = 4.19$, $p < .02$). When controlling for the impact of AFE, the task statistically significantly impacted the BESA Index score: ($\beta = .31$, $t(46) = 2.01$, $p = .05$). The standardized coefficient of .31 can be characterized as a medium effect size. The coefficient can be interpreted to mean that a 1 SD increase in the LN task performance would lead to a .31 SD increase on the Spanish BESA Index score. For the QU task, the overall impact of AFE and the QU task is statistically significant ($F[2, 44] = 4.47$, $p = .02$). The QU task impact, when controlling for AFE, is statistically significantly related to the BESA Spanish Index ($\beta = .31$, $t(46) = 2.14$, $p = .04$). The standardized coefficient showed that an increase in 1 SD increase on the QU task would lead to a .31 SD increase on the BESA score, an effect size that can be characterized as medium. Finally, for the CTOPP task, the overall relationship is statistically significant ($F[2, 44] = 3.47$, $p = .04$). However, when controlling for AFE, the relationship between the CTOPP and the BESA Spanish Index is not statistically significant ($\beta = .25$, $t(46) = 1.65$, $p = .11$). Therefore, in contrast to the English BESA Index, for which all NWR tasks significantly predicted the BESA Index score, when controlling for AFE, only the LN and the QU tasks are significantly related to the Spanish Index score on the BESA.

For the BESA Index, which uses the stronger score, regardless of language, it was predicted that all three NWR tasks would impact the BESA score. This prediction was supported by the regression results. Using a simple regression, each NWR task was regressed on the BESA Index score. All of the NWR tasks were significantly related to the BESA score, when controlling for age of first exposure. For the LN task, the NWR task was statistically significantly related to the BESA Index ($F[2, 44] = 21.64$, $p < .001$). When controlling for the impact of AFE, the LN task statistically significantly impacted the BESA Index sore: *($\beta = .67$, $t(46) = 5.63$, $p < .001$)*. The standardized coefficient of .67 can be characterized as a large effect size, indicating that a 1 SD increase in the LN task performance would lead to an .67 SD increase on the BESA Index score. For the QU task, the overall impact of AFE and the QU task is statistically significant ($F[2, 44] = 12.23$, $p < .001$). The QU task impact, when controlling for AFE, is statistically significant *($\beta = .50$, $t(46) = 3.92$, $p < .001$)*. The standardized coefficient showed that an increase of 1 SD on the QU task would lead to a .5 SD increase on the BESA score, an effect size that can be characterized as large. Finally, for the CTOPP task, the overall relationship is statistically significant ($F[2, 44] = 18.24$, $p < .001$). For the CTOPP task performance, when controlling for AFE, the relationship is statistically significant *($\beta = .61$, $t(46) = 5.08$, $p < .001$)*. A 1 SD increase on the CTOPP leads to a .61 SD increase on the BESA, a large effect.

Similar to the English BESA Index, using the overall BESA Index revealed that all NWR tasks significantly predicted the BESA Index score, when controlling for AFE. As a reminder, a low score on the BESA is indicative of language impairment in bilingual children. In other words, the two tasks designed for use with bilingual children are predictive of both their Spanish and English ability, whereas the CTOPP is only predictive of their English ability.

In order to examine whether these relationships are constant across the BESA scores, a quantile regression was conducted for the BESA Index score, which is indicative of language impairment. Analyses were conducted using the 16^{th}, 25^{th}, and 50^{th} quantile, which correspond to standard clinical cutoffs. The main analysis determines the intercept and slope for each quantile. A test of equality of slopes indicates whether the parameters for the quantiles differ statistically. Follow-up testing, using ANOVA, determines which combination of quantiles has statistically significantly different parameters. It was predicted that the parameters from lower quantiles would significantly differ from those of the higher quantiles for the QU and LN tasks, indicating that the NWR tasks were more strongly related to the BESA score at those lower quantiles than at the upper quantiles. Initial quantile regression analyses were performed in SPSS. For a more direct comparison, a regular regression was estimated, without AFE as a covariate. In addition, for the following analyses, the independent variables (i.e., the NWR task scores) were centered at the mean in order to increase the interpretability of the regression intercepts.

For the LN task, the test of Equality of slopes revealed that slopes at each percentile were significantly different from each other (F[2, 139) = 4.42, p = .01). For the QU task, the test of Equality of slopes revealed that slopes at each percentile were significantly different from each other (F[2, 139] = 3.41, p = .04). Finally, for the CTOPP, the test of Equality of slopes revealed that slopes at each percentile were significantly different from each other (F[2, 139] = 3.71, p = .03). These quantile regression results are presented in Figures 2-4. In these figures, the x-axis represents the quantiles examined, while the y-axis represents the value of the estimates.

Whereas the dotted red lines represent the confidence interval for the OLS regression, the solid red line shows the OLS regression line. Visually inspecting the graphs for all of the NWR tasks illustrates that the lower quantiles have a slope value that is outside of the confidence interval of the slope of the OLS regression line, indicating that there is a significant difference between the quantile regression and the OLS regression at these quantiles. The Individual Tests of Equality of Slopes revealed a statistically significant difference for all three tasks, but it is not clear at which quantiles the difference lies.

Follow-up analyses, using an ANOVA, were conducted using the qauntreg package in R (R Core Team, 2015) using procedures described in Koenker, 2015. These comparisons analyze whether slope coefficients for each predictor were significantly different between quantiles. The comparisons were conducted between the 16^{th}, 25^{th}, and 50^{th} quantiles. For the LN task, the 16^{th} percentile differed statistically from the 50^{th} quantile ($F[1,93] = 7.03$, $p = .01$). In addition, the 25^{th} quantile differed significantly from the 50^{th} ($F[1,93] = 5.30$, $p = .02$). For the QU task, a significant difference was found between the 16^{th} and 50^{th} quantile ($F[1,93] = 5.70$, $p = .02$). For the CTOPP task, the overall slopes analysis was significant. Significant differences were found between the 16^{th} and 50th quantile ($F[1,93] = 4.32$, $p = .04$). The difference between the 25^{th} and 50^{th} percentile approached significance ($F[1,93] = 3.67$, $p = .058$). Thus, the quantile comparison tests revealed several significant differences between slope coefficients.

A comparison between the OLS regression results and the quantile regression results is outlined in Table 14. In addition, Figures 5-7 show the regression slopes for the OLS and quantile regression for each task. The table lists the intercepts as well as slopes for the regression lines. For all of the NWR tasks, the slope coefficients are larger at the 16^{th} percentile and then decrease across the 25^{th} and 50^{th} quantile. For all three tasks, the difference between the

16th and 50th quantile is statistically significant, indicating that the relationship between the NWR task and the BESA is greater at the 16th than the 50th quantile. In other words, while there is an overall relationship between the NWR tasks and the BESA, as shown by the OLS regression, this relationship is stronger at the lower BESA score. This suggests that all three NWR tasks have the potential to be used as a diagnostic tool for language impairment in bilingual children.

Chapter 5: Discussion

This study examined the diagnostic potential of three nonword repetition (NWR) tasks that differed in design to determine how bilingual children perform on these tasks and whether performance is influenced by a child's socioeconomic status (SES), age of first exposure (AFE), or current language use. Finally, this study determined to what extent each NWR task predicts a child's language ability. The purpose of this study was to determine which NWR task has the most potential for use as a diagnostic tool for identifying whether a bilingual child has SLI. In comparing the three tasks, the goal was to determine which task minimizes the effects of environmental factors such as SES or language experience and maximizes the difference between typically developing children and those with language impairment. A series of regressions (SES by task performance, AFE by task performance, and NWR by the language score) examined the answers to the research questions. A quantile regression determined whether the relationship between the NWR task and the language assessment differed across quantiles, thus providing a more nuanced look at the participants' scores on the NWR tasks and how they relate to their language scores on the Bilingual English Spanish Assessment (BESA).

Summary and Interpretation of Findings

The data suggest that all of the findings of this study may have been impacted by the participants' dominance in English. Since all of the data were collected in the United States, and almost all of the children attended English-language schools, the children's English scores carried most of the predictive power for NWR performance (and perhaps language impairment). Recall that the children's English scores (morphology and semantics) were significantly higher than their scores in Spanish. All of the results should be interpreted in light of this language dominance, with the understanding that children dominant in Spanish many perform differently than the participants of this study.

For the first research question on children's pattern of performance on the three differently designed NWR tasks, all of the scores decrease as item length increases, a result that corresponds to ample research showing the same pattern. Therefore, all three NWR tasks examined in this study are measures of phonological memory for bilingual children. The QU task may function slightly differently from the other two tasks, based on the fact that for all syllable lengths, participants scored higher on this task than the other two. The reason for this difference in the QU task was not immediately apparent. Examining the task items indicated that while the task included later-developing sounds in English, such as /l/ and included a sound not found in Spanish, /z/, the task also included four items with real words in English and Spanish. This inclusion of real words may have introduced this task advantage. Table 10 outlines details on the task items. Children scored best on the QQ task and lowest on the CTOPP overall, with scores on the LN task falling in between those of the other two tasks. A future study could examine performance on individual items, as an item analysis, to determine which items were easier or harder for children to repeat.

For the research question on whether SES influences the nonword repetition scores, the data illustrated that SES, measured as parental education in years, did not affect task performance, mirroring findings by Engel, Santos, and Gathercole (2008) who also determined that there was no difference between low and high SES participants' scores on a NWR task. This provides support for the idea that NWR tasks are less biased than other, knowledge-based tasks and are thus instruments that are appropriate for use with culturally and linguistically diverse groups (Campbell, Dollaghan, Needleman, and Janosky, 1997; Dollaghan, Campbell, Needleman, and Dunlosky, 1997). This is particularly important for studies including bilingual children, who are more likely to come from lower SES families (Caldéron, 2003). If indeed NWR tasks are good indicators of language impairment, the fact that SES does not affect task performance is an additional argument for their use as part of an assessment protocol. These results provide further support for the use of the three NWR tasks in assessments of bilingual children.

For the question of whether age of first exposure (years from 0-6) impacts NWR performance, testing showed that age of first exposure (AFE) influenced the LN task but not the other two tasks. This means that children's performance on the LN task may be affected by the age at which they began learning English. The relationship between AFE and the performance on the LN task was negative, with an earlier AFE corresponding to higher task performance. In other words, a child who began learning English at birth may outperform a child who began learning English later, regardless of the children's phonological memory skills. The role of AFE was not the focus of the current study but could be explored in future research involving the three NWR tasks.

Once AFE was controlled for in the statistical analyses, the research question exploring the predictive power of percentage of current English use (input and output, averaged) and performance on the three NWR tasks was addressed. The role of the children's dominant language (English) again played a part. Children's current use of English may be predictive of their performance on the LN and the CTOPP, as these regression analyses approached significance. Further research should include a larger number of participants in order to increase the power of statistical analyses. For both tasks, the effect size was large, suggesting that these analyses would have been significant with an increase in the number of participants. Unlike the

LN and the CTOPP, the QU task was less influenced by current English use when controlling for AFE. For this task, the effect size was medium-to-large, indicating a smaller effect than for the other two tasks. This analysis lends further support to the idea that the QU task differs from the other two tasks, as revealed in the overall task performance.

Prior research on the role of language experience in NWR task performance has been inconclusive, with some studies providing evidence for the impact of language experience, and other studies showing no difference. For example, Thordardottir and Juliusdottir (2013) found that children's score on a NWR task based in the L2 was similar to a task based in the L1. Similarly, Lee and Gorman (2012) determined that bilingual children scored similarly to monolingual children on a NWR task based in the L2. Finally, Thordrdottir and Brandeker (2013) found that language experience had little impact on NWR task performance. Much previous research examined NWR performance in light of current language use, rather than cumulative use over a child's lifetime. One exception was Summers, Bohman, Gillam, Peña, and Bedore (2010), who determined that a child's cumulative language exposure had a greater influence on NWR task performance than current language use. In other words, the current usage of English may not be the most accurate predictor of performance. Rather, the cumulative language use over a child's lifespan would have a greater effect. Future research should compare current use with cumulative use to determine whether the impact on NWR performance differs for these three tasks.

The results of the current study indicate that the role of language experience depends on task design. The QU task, which was designed to minimize the effect of language-specific characteristics, was less influenced by current English usage than the other two tasks. This is surprising, given the fact that the LN task was designed with complete overlap of phonological characteristics between the two languages. This task relates to the prior research that has shown that the phonological inventory of a child's languages influences task performance (Lee and Gorman, 2013; Sharp and Gathercole, 2013). Some unintended underlying task characteristics may have influenced task performance. As mentioned earlier, a follow-up analysis could be conducted, with an item analysis to determine if certain task items influenced overall task performance. Alternately, an individual participant analysis could be conducted to determine whether individual participants impacted overall results.

Because the results for the analysis of language experience and NWR tasks were unexpected, follow-up analyses examined the role of articulation in task performance. Again the QU task appears to function differently from the other tasks. While the Spanish phonology task, which assesses articulation skills, significantly related to the LN task performance, the English phonology subtest was statistically significantly related to the CTOPP. This suggests that some of the features of the LN task may make it susceptible to the influence of a child's articulatory skills in Spanish. This implies that the construction of the QU task was indeed "universal," with a child's phonological skills in either language not influencing their performance on this task. Thus, this task may be a more accurate measure of phonological memory for bilingual children than the other two tasks. The pattern of results indicated that the QU task may be a better candidate than the other two NWR tasks for inclusion in a battery of tests to diagnose Specific Language Impairment. As stated earlier, the results regarding the phonology subtest on the BESA and the NWR task performance are in line with previous research findings that show that bilingual children's performance on NWR tasks is influenced by the phonology of the task items. Because children draw on their experience with the phonology of their languages, NWR task performance improves when the nonwords are constructed using phonological repertoires and phonotactic rules of the native language of the child, assuming that this is the dominant language (Masoura & Gathercole, 1999; Thorn & Gathercole, 1999).

When further examining the impact of phonology, a developmental scoring method was used to control for participants' articulation errors. This scoring method did not change the overall picture of results, indicating that children's performance on the NWR tasks truly measures phonological memory rather than articulatory skill. In the literature regarding bilingual children's acquisition of speech sounds, sound that are common to both of the child's languages may be acquired more quickly by bilingual children than monolingual children (Fabiano-Smith & Barlow, 2010; Gildersleeve-Neumann & Wright, 2010). While the acquisition of speech sounds found only in one of the child's two languages may take longer than for monolingual children, by the age of four, these differences in phonological development largely disappear and typically developing bilingual children performed commensurate with their monolingual peers (Goldstein & Washington, 2001; Goldstein, Fabiano, & Washington, 2005). Therefore, since the participants in this study were all at least five years of age, their articulatory skill in both languages should be commensurate and should not influence task performance.

The third research question asked about each task's predictive ability in identifying children with language scores at the 16^{th}, 25^{th}, and 50^{th} quantile of the BESA Index score. Before answering this question, the relationship between each NWR task and the BESA test was examined using a standard linear regression. These analyses revealed that all of the NWR tasks were related to the English BESA Index score, when controlling for AFE. In addition, the LN and QU task were related to the Spanish BESA Index score. This pattern was not found for the relationship between the CTOPP and the Spanish BESA Index. This is logical given the fact that the CTOPP was designed for use with monolingual English speakers and would therefore not be predictive of Spanish language ability. The CTOPP includes speech sounds not found in Spanish and includes complex syllable structures that are not common in Spanish. It is possible that if the participant group was dominant in Spanish, the CTOPP would be an unreliable measure in predicting the participant's overall language ability, across languages. However, the LN and QU task would likely continue to relate to children's language scores if they are dominant in Spanish. This possibility needs to be assessed in future studies including more participants dominant in Spanish.

All three NWR tasks were predictive of children's overall BESA score, which was the standardized measure used as a diagnostic tool for language impairment, suggesting that all three tasks have potential as a diagnostic marker for language impairment. However, the pattern of predictions is the same for the English BESA Index and the overall index, due to the fact that the English score carries the overall BESA Index score. As stated earlier, this result reflected the participants' English language dominance. These results are in line with previous research findings that have shown that bilingual children's performance on NWR tasks is influenced by vocabulary knowledge in the language on which the task is based (Ebert, Pham, & Kohnert, 2014; Lee & Gorman, 2012; Parra, Hoff, & Core, 2011; Peña, Bedore, & Zlatic-Giunta, 2002; Summers et al., 2010; Thorn & Gathercole, 1999). In the current study, the CTOPP is the only task based on a particular language (English), and this task is statistically related to knowledge of the English language, as evidenced by participants' performance on the BESA English Index. In other words, this task may be a good diagnostic marker for language impairment in bilingual children who are dominant in English. However, the CTOPP would be less accurate for children who are dominant in Spanish, as evidenced by the lack of a significant relationship between the CTOPP and the Spanish language BESA score. The other two tasks, which are statistically related to the BESA English and Spanish score, have more potential as a diagnostic tool for children with language dominance in either their L1 or L2. Of the two tasks, the QU is the stronger candidate, due to the lesser influence of language experience on task performance. Thus, a child's language experience impacts the strength of the relationship between the NWR task and the BESA.

To examine each task's predictive power for language impairment, an additional method of analysis was used. This analysis answered the third research question, asking whether the relationship between each NWR task and the BESA score changed at different quantiles of the BESA. In other words, quantile regression allows the researcher to answer not only the question "is there a relationship between the NWR tasks and the BESA score?" but also the question "does this relationship change across different scores on the BESA test?" The pattern found here indicated that the relationship between all three NWR tasks was stronger for lower quantiles than higher quantiles. In other words, a lower performance on the BESA was more strongly related to NWR performance, indicating that all three NWR tasks have the potential for use as part of a testing protocol for language impairment. The strength (or slope) of the relation between the NWR and BESA decreased as the quantile of BESA increased. While the overall conclusion regarding the tasks' potential as a diagnostic marker did not change in the current study, quantile regression allowed a more nuanced examination of the data. Since the regular OLS regression results were statistically significant, indicating a relationship between all three NWR tasks and the BESA scores, the quantile regression did not add any information in terms of clinical application of the tasks. However, if the quantile regression results had shown a different pattern, the interpretation of the OLS regression would have changed. Thus, in the current study, the quantile regression confirmed the results of the OLS regression rather than changing the interpretation of the OLS results. While not adding to the clinical implications of study findings, the current study illustrated the potential for quantile regression in research on diagnostic accuracy. Quantile regression has several advantages, such as using all data points, eliminating the need for pre-identified groups of children with and without language impairment. In future studies, quantile regression analyses should be used when employing a one-gate design, which does not separate participants into diagnostic groups. Utilizing quantile regression ensures that an OLS regression does not misrepresent the relationship between variables, which may change dependent on the participant scores on the measures.

In summary, when examining the three tasks, the CTOPP was the most biased task in terms of language experience and language knowledge. Task performance was influenced by both the amount of English spoken/heard by the child and the child's knowledge of English vocabulary and syntax. Therefore, this task has limited utility as a diagnostic tool, especially for children who have more experience with and knowledge of Spanish. The LN task fared somewhat better but still exhibited a possible impact of language experience. The QU task fared best because it was not impacted by language experience, age of first exposure, or articulation skills. Furthermore, the task was related to language skills in both Spanish and English. Thus, of the three NWR tasks examined, the QU task shows the most potential for use as a diagnostic tool for language impairment in bilingual children.

Study Significance

This study has addressed several shortcomings of most previous studies on the diagnostic accuracy of NWR tasks as a tool to identify bilingual children with SLI. This study has improved the research design, systematically manipulated the design of NWR tasks, and examined two scoring methods of NWR tasks. The current study provides details on the children's acquisition status (sequential vs. simultaneous) by coding for participants' age of first exposure and quantifying current language experience. Much of the prior research did not include details on children's language status and merely reported that children were dominant in one language without providing detailed percentages of use. The current study included participants within a narrow age range, increasing the ability to compare the results of this study to those of prior studies conducted with monolingual children with SLI. It has added to the development of NWR tasks to be used with bilingual children in that the tasks used in this study were designed to manipulate the amount of overlap between the task phonemes and those known by a bilingual child. While one NWR task was based solely in English and one was based on the overlap between Spanish and English, the third task was designed to be "universal" across most of the world's languages. The evidence supports the use of this universal task for bilingual children because this task is not influenced by current English usage, is significantly related to language scores in both Spanish and English, and shows a stronger relationship between task performance and a low BESA score than a higher BESA score. The QU task differs from the CTOPP and LN task, which may both be influenced by current English usage. While the LN task is significantly related to language scores in English and Spanish, the CTOPP task is only related to English scores. Therefore, the QU task meets all of the requirements for being a good indicator of language impairment in bilingual children, whereas the other two tasks only meet some of the requirements.

This study's scoring method allowed for articulation errors based on a child's L1, including those based on a child's dialect. Further testing was completed to determine whether children's articulation errors significantly impacted their NWR performance. This study revealed that articulation did not change the interpretation of the NWR tasks and how they relate to a child's language ability. Therefore, articulation did not make a difference in the production of the nonwords for bilingual children, as long as those children are of an age at which articulatory skills should be commensurate with those of monolingual children.

This study uses research methods not frequently utilized in diagnostic accuracy research: through its use of participant recruitment and its use of quantile regression as an analysis technique. This study sampled participants from the community using a snowball technique. While not a truly random sample, this method allowed the researchers to remain blinded to the participants' language status and is thus a one-gate design. Two separate researchers tested children in their two languages; thus, blinding included both the children's status regarding language impairment as well as language dominance. Finally, all participants received all of the same testing measures. This ensured that diagnostic accuracy did not result in false negatives through the use of a pre-identified impaired group.

The statistical method of quantile regression also adds to the nuance of interpretation of NWR task results. This relatively newly applied statistical procedure may reveal a detailed pattern of relationships among variables that would be missed in a more traditional regression approach. Quantile regression avoids the loss of power when creating subgroups (those with the disease versus those without). While the OLS regression illustrated a statistically significant relationship between the three NWR tasks and the BESA score, the quantile regression allowed a detailed interpretation of the results across test scores. For all NWR tasks, the relationship between the NWR task and the BESA score was stronger at the 16th percentile than at the higher percentiles, indicating that a low score on the BESA was more strongly related to a low score on the NWR task than a higher BESA score. Quantile regression is a new methodology in diagnostic research and has none of the issues related to having an arbitrary cutoff point, such as that used when conducting sensitivity and specificity testing. Future research on the use of NWR tasks in diagnosing language impairment in bilingual children should use quantile regression analyses to address some of the challenges of research with this heterogeneous population.

Clinical Implications

All three of the NWR tasks tested in this study have clinical potential and can address the difficulty of assessing bilingual children, who vary on numerous characteristics, such as age of first exposure or amount and quality of input in their languages. NWR tasks may be administered by a monolingual Speech-Language Pathologist, are easy to administer, and take few resources, such as expensive standardized assessments in both of a child's languages. All three tasks examined in this study are predictive of children's language ability, provided that children are dominant in English rather than their first language. The implication is, as long as a child's language dominance is determined to be English (based on parent and teacher report), any of these three tasks would be informative. However, clinicians should be mindful of the influence of language experience. Both the LN and the CTOPP may be influenced by the amount of English a child currently hears and speaks. Therefore, these tasks may not be clinically informative for children whose current English exposure is lower or whose current exposure to their two languages is more balanced. According to this study, the QU task has more clinical potential because it was not influenced by a child's current use of English. While it is not a norm-referenced task, research on its clinical use is ongoing and normative studies are currently being conducted. On a more theoretical level, clinicians who assess bilingual children for language impairment should keep in mind that factors such as task design (length of items, phonemes that make up items, phonetic realization of phonemes, etc.) may impact bilingual children's performance on the task, leading to a misrepresentation of a child's phonological memory.

Study Limitations and Future Directions

This study is somewhat limited by the language backgrounds of its participants. The study's participants were dominant in English, which is a result of the geographic location of study recruitment. Because of this, any conclusions drawn may not be generalizable to participants who are dominant in Spanish. In order to address the possible impact of language dominance, this study should be repeated in a Spanish-dominant environment to assess whether results would differ in that context. Another possibility is to conduct follow-up analyses of the data collected for the current study, by separating children with less exposure to English. It would be possible to re-run the analyses with this language group to determine whether the statistical results change. For children who are dominant in Spanish, it is expected that the CTOPP task would not relate to language ability, as measured on the BESA. Further, while children in this study varied on current language exposure, with a range of 20-80% exposure to English, it is possible that cumulative language exposure would have an impact on task performance. This possibility could be explored systematically and findings could be compared to those of Summers et al. (2010). Finally, an item analysis should be conducted for the three NWR tasks. This would allow the researcher to delve into possible design features that may have impacted performance, such as syllable structure, speech sounds used, or aspects of the production and recording of sounds. An item difficulty analysis could be completed in order to identify specific items that may have impacted participant performance. These items could be eliminated or altered, in the case of the QU and the LN task, and testing could be repeated with a revised version of the tasks. In terms of the CTOPP, identifying problematic items may help an SLP alter their scoring method or interpretation of student performance on problematic items.

Future research should focus on the QU task, which showed the most clinical potential, and should be expanded to participants from other language groups. This task has the distinct advantage of being "universal," thus requiring no adaptations for use with other populations. This task can thus be examined in the same manner as the current study, in terms of language experience and language ability, with different L1 groups.

Conclusion

The current study adds to the field of assessment of bilingual children by examining children's performance on three different nonword repetition (NWR) tasks as related to their language experience and language ability. Because this study includes bilingual children in a narrow age range that corresponds to the literature on SLI in monolingual children, it makes comparisons between studies easier. In addition, the current study systematically examined three different NWR tasks, exploring differences between task performance and how it relates to language experience and ability. Thus, this study adds to the theoretical knowledge on NWR tasks and how bilingual children perform on these tasks. Finally, this study used a sampling method that utilized a one-gate design, without prior determination of typically developing children and those with language impairment. This eliminated the bias inherent in using pre-diagnosed children as a reference group. The results of this study enable monolingual Speech Pathologists to determine task utility in assessing bilingual children for SLI. Results for the proposed study represent a significant step forward in improving the assessment of children with SLI, and will thus have clinical and theoretical importance. These findings will be informative for clinical practice as well as theoretical knowledge because they provide knowledge on the overall pattern of performance as well as information used in clinical decision making.

References

Archibald, L. M. (2008). The promise of nonword repetition as a clinical tool. *Canadian Journal of Speech-Language Pathology and Audiology, 32*(1), 21-28.

Archibald, L. M., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders, 41*(6), 675-693.

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

Bishop, D. V. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language & Communication Disorders, 49*(4), 381-415.

Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of child Psychology and Psychiatry, 37*(4), 391-403.

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A Quasi-Universal Nonword Repetition Task as a Diagnostic Tool for Bilingual Children learning Dutch as a Second Language. *Journal of Speech, Language, and Hearing Research*.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *GlotInternational* 5:9/10, 341-345.

Brandeker, M., & Thordardottir, E. (2015). Language exposure in bilingual toddlers:

performance on nonword repetition and lexical tasks. *Am J Speech Lang Pathol, 24*(2), 126-138. doi:10.1044/2015_AJSLP-13-0106

Bricker, D., Squires, J., Mounts, L., Potter, L., Nickel, R., Twombly, E., & Farrell, J. (1999). Ages and stages questionnaires. *Baltimore, MD: Paul H. Brookes*.

Calderon, J., & Gutierrez-Clellen, V. (2003). Nonword repetition in emerging bilingual children with language impairment. In *4th International Symposium on Bilingualism, Tempe, AZ.*

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing dependant measures. *Journal of Speech, Language, and Hearing Research, 40*(3), 519-525.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*(6), 1142-1157.

Chiat, S. (2015). Non-Word Repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment* (Vol. 13, pp. 125). Bristol: Multilingual Matters.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders, 43*(1), 1-40.

Conti-Ramsden, G., & Windfuhr, K. (2002). Productivity with word order and morphology: A comparative look at children with SLI and children with normal language abilities. *International Journal of Language & Communication Disorders, 37*(1), 17-30.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of child Psychology and Psychiatry, 42*(6), 741-748.

Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-word and nonword repetition in Italian-speaking children with specific language impairment: a study of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 56*(1), 323-336.

Dollaghan, & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41*(5), 1136-1146.

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

Dollaghan, C., Campbell, T., Needleman, H., & Dunlosky, G. (1997). Reducing bias in language assessment: A processing-dependent measure. *Journal of Speech, Language and Hearing Research, 40*, 519-525.

Ebert, K. D., Kalanek, J., Cordero, K. N., & Kohnert, K. (2008). Spanish nonword repetition stimuli development and preliminary results. *Communication Disorders Quarterly, 29*(2), 67-74.

Ebert, K. D., Pham, G., & Kohnert, K. (2014). Lexical profiles of bilingual children with primary language impairment. *Biling (Camb Engl), 17*(4), 766-783. doi:10.1017/S1366728913000825

Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research, 47*(2), 421-436.

Engel de Abreu, P. M. J., Gathercole, S. E., & Martin, R. (2011). Disentangling the relationship between working memory and language: The roles of short-term storage and cognitive control. *Learning and Individual Differences, 21*(5), 569-574. doi:10.1016/j.lindif.2011.06.002

Engel, P. M. J., Santos, F. H., & Gathercole, S. E. (2008). Are Working Memory Measures Free of Socioeconomic Influence? *Journal of Speech, Language & Hearing Research, 51*(6), 1580-1587.

Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 50*(1), 177-195.

Fabiano-Smith, L., & Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism*, *13*(1), 81-97.

Fey, M. E. (1992). Articulation and PhonologyInextricable Constructs in Speech Pathology. *Language, Speech, and Hearing Services in Schools*, *23*(3), 225-232.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*(3), 336-360.

Gildersleeve, C., Davis, B., & Stubbe, E. (1996). *When monolingual rules don't apply: Speech development in a bilingual environment.* Paper presented at the annual convention of the American Speech-Language-Hearing Association, Seattle, WA.

Gildersleeve-Neumann, C., & Davis, B. (1998). *Learning English in a bilingual preschool environment: Change over time.* Paper presented at the annual convention of the American Speech-Language-Hearing Association, San Antonio, TX.

Gildersleeve-Neumann, C. E., & Wright, K. L. (2010). English speech acquisition in 3-to 5-year-old children learning Russian and English. *Language, Speech, and Hearing Services in Schools*, *41*(4), 429-444.

Gillam, R. B., Peña, E. D., Bedore, L. M., Bohman, T. M., & Mendez-Perez, A. (2013). Identification of specific language impairment in bilingual children: I. Assessment in English. *Journal of Speech, Language, and Hearing Research, 56*(6), 1813-1823.

Girbau, D., & Schwartz, R. G. (2008). Phonological working memory in Spanish-English bilingual children with and without specific language impairment. *J Commun Disord, 41*(2), 124-145. doi:10.1016/j.jcomdis.2007.07.001

Goldstein, B. (2007). Spanish speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 539-553). Clifton Park, NY: Thomson Delmar.

Goldstein, B., & Washington, P. S. (2001). An initial investigation of phonological patterns in typically developing 4-year-old Spanish-English bilingual children. *Language, speech, and hearing services in schools, 32*(3), 153-164.

Goldstein, B. A., Fabiano, L., & Washington, P. S. (2005). Phonological skills in predominantly English-speaking, predominantly Spanish-speaking, and Spanish-English bilingual children. *Language, speech, and hearing services in schools, 36*(3), 201-218.

Grimm, A., & Schulz, P. (2014). Specific Language Impairment and Early Second Language Acquisition: The Risk of Over- and Underdiagnosis. *Child Indicators Research, 7*(4), 821-841. doi:10.1007/s12187-013-9230-6

Guiberson, M., & Rodriguez, 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.

Guiberson, M., Rodríguez, B. L., & Dale, P. S. (2011). Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Language, speech, and hearing services in schools, 42*(4), 536-549.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24*(02), 267-288.

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.

Hansson, K., Nettelbladt, U., & Leonard, L. B. (2000). Specific Language Impairment in SwedishThe Status of Verb Morphology and Word Order. *Journal of Speech, Language, and Hearing Research, 43*(4), 848-864.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child development*, *74*(5), 1368-1378.

Huls, S., Storkel, H., and Schuh, N. (2015, September). What's in a (Non)word: Developing a "language neutral" nonword repetition task for use with bilingual students Poster session presented at the Annual Convention of the Kansas Speech Language and Hearing Association, Wichita, KS.

Jackson-Maldonado, D., Bates, E., & Thal, D. (1992). Fundación MacArthur: Inventario del desarrollo de habilidades comunicativas. *San Diego, CA: San Diego State University*. Jakubowicz, C., Nash, L., Rigaut, C., & Gerard, C.-L. (1998). Determiners and clitic pronouns in French-speaking children with SLI. *Language acquisition, 7*(2-4), 113-160.

Jarvis, B. G. (2014). MediaLab (Version 2014.1.123) [Computer Software]. New York, NY: Empirisoft Corporation.

Koenker, R. (2015). quantreg: Quantile Regression. R package version 5.11. http. *CRAN. R-project. org/package= quantreg*.

Kohnert, K. (2013). *Language disorders in bilingual children and adults*: Plural Publishing.

Kohnert, K., Windsor, J., & Ebert, K. D. (2009). Primary or "specific" language impairment and children learning a second language. *Brain Lang, 109*(2-3), 101-111. doi:10.1016/j.bandl.2008.01.009

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research & Practice, 21*(1), 19-29.

Lee, & Gorman, B. K. (2012). Nonword repetition performance and related factors in children representing four linguistic groups. *International Journal of Bilingualism, 17*(4), 479-495. doi:10.1177/1367006912438303

Lee, Kim, Y. T., & Yim, D. (2013). Non-word repetition performance in Korean-English bilingual children. *International journal of speech-language pathology, 15*(4), 375-382.

Leonard, L. B. (2014). *Children with specific language impairment*: MIT press.

Linden, M. v. d. (2009). *Language assessment in multilingual children: the Non-word Repetition Task.* (Masters), Utrecht University, Utrecht. (3035433)

Mancilla-Martinez, J., Gámez, P. B., Vagh, S. B., & Lesaux, N. K. (2015). Parent Reports of Young Spanish–English Bilingual Children's Productive Vocabulary: A Development and Validation Study. *Language, speech, and hearing services in schools*, 1-15.

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign language learning. *International Journal of Psychology, 34*(5-6), 383-388.

McLeod, S. (2007). *The international guide to speech acquisition*: Thomson Delmar Learning Clifton Park, NY.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 48*(5), 1033-1047.

Paradis, J. (2015). The development of English as a second language with and without specific language impairment: Clinical implications. *Journal of Speech, Language, and Hearing Research*.

Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: using parent report on first language development. *J Commun Disord, 43*(6), 474-497. doi:10.1016/j.jcomdis.2010.01.002

Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development and disorders: A handbook on bilingualism and second language learning*: ERIC.

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

Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *J Exp Child Psychol, 108*(1), 113-125. doi:10.1016/j.jecp.2010.07.011

Peña, E. D., Bedore, L. M., & Zlatic-Giunta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech, Language, and Hearing Research, 45*(5), 938-947.

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B., & Bedore, L. M. (2014). *BESA: Bilingual English-Spanish Assessment Manual*: AR-Clinical publications.

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). <Poll et al. 2010.pdf>. *Journal of Speech, Language & Hearing Research, 53*, 414-429.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org.

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

Reynolds, C. R., & Kamphaus, R. (2009). Development and application of the Reynolds Intellectual Assessment Scales (RIAS). *Practitioner's Guide to Assessing Intelligence and Achievement*, 95.

Rowe, M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of child language*, *35*(01), 185-205.

Roy, P., & Chiat, S. (2004). A Prosodically Controlled Word and Nonword Repetition Task for 2- to 4-Year-Olds: Evidence From Typically Developing Children. *Journal of Speech, Language, and Hearing Research, 47*(1), 223-234.

Roy, P., Chiat, S., & Marshall, C. R. (2013). Teasing apart disadvantage from disorder. *Current issues in developmental disorders*, 125-150.

Sharp, K. M., & Gathercole, V. C. M. (2013). Can a novel word repetition task be a language-neutral assessment tool? Evidence from Welsh–English bilingual children. *Child Language Teaching and Therapy*, *29*(1), 77-89.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders, 55*(4), 779-798.

Summers, C., Bohman, T. M., Gillam, R. B., Peña, E. D., & Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *Int J Lang Commun Disord, 45*(4), 480-493. doi:10.3109/13682820903198058

Theodorou, E., & Grohmann, K. K. (2015). Object clitics in Cypriot Greek children with SLI. *Lingua*.

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *J Commun Disord, 46*(1), 1-16. doi:10.1016/j.jcomdis.2012.08.002

Thorn, A. S., & Gathercole, S. E. (1999). Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *Q J Exp Psychol A, 52*(2), 303-324. doi:10.1080/713755823

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 39*(6), 1284-1294.

Tomblin, J. B., Smith, E., & Zhang, X. (1997). Epidemiology of specific language impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders, 30*(4), 325-344. doi:http://dx.doi.org/10.1016/S0021-9924(97)00015-4

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and speech, 40*(1), 47-62.

Wagner, R., Torgesen, J., & Rashotte, C. (1999). Comprehensive test of phonological processing (CTOPP). *Pro-Ed, Austin, Texas*.

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.

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2012). Preschool Language Scale, (PLS-4) Spanish Edition. *Age*.

Figures

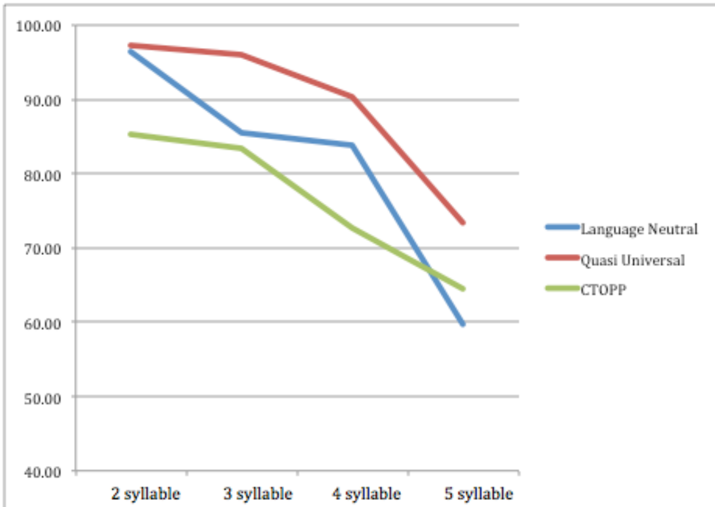

Figure 1: Nonword Repetition Task Performance

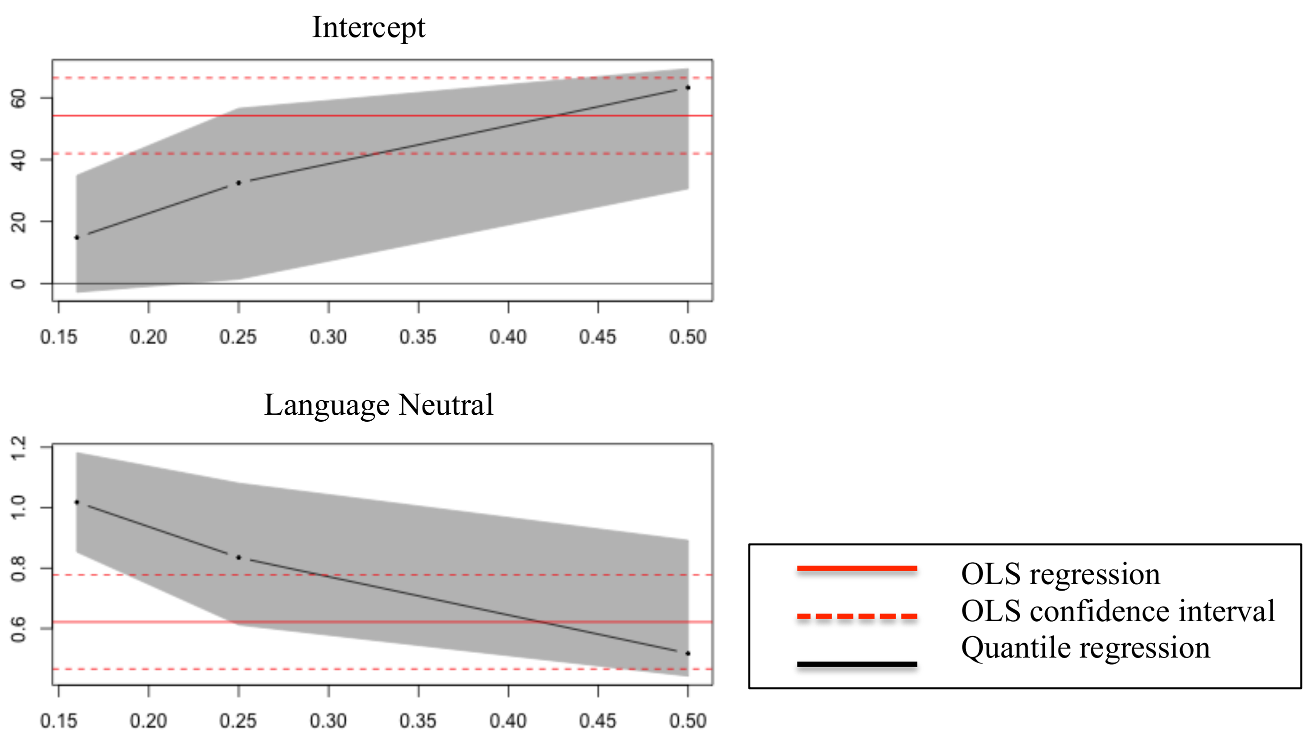

Figure 2: Quantile Regression for Language Neutral Task

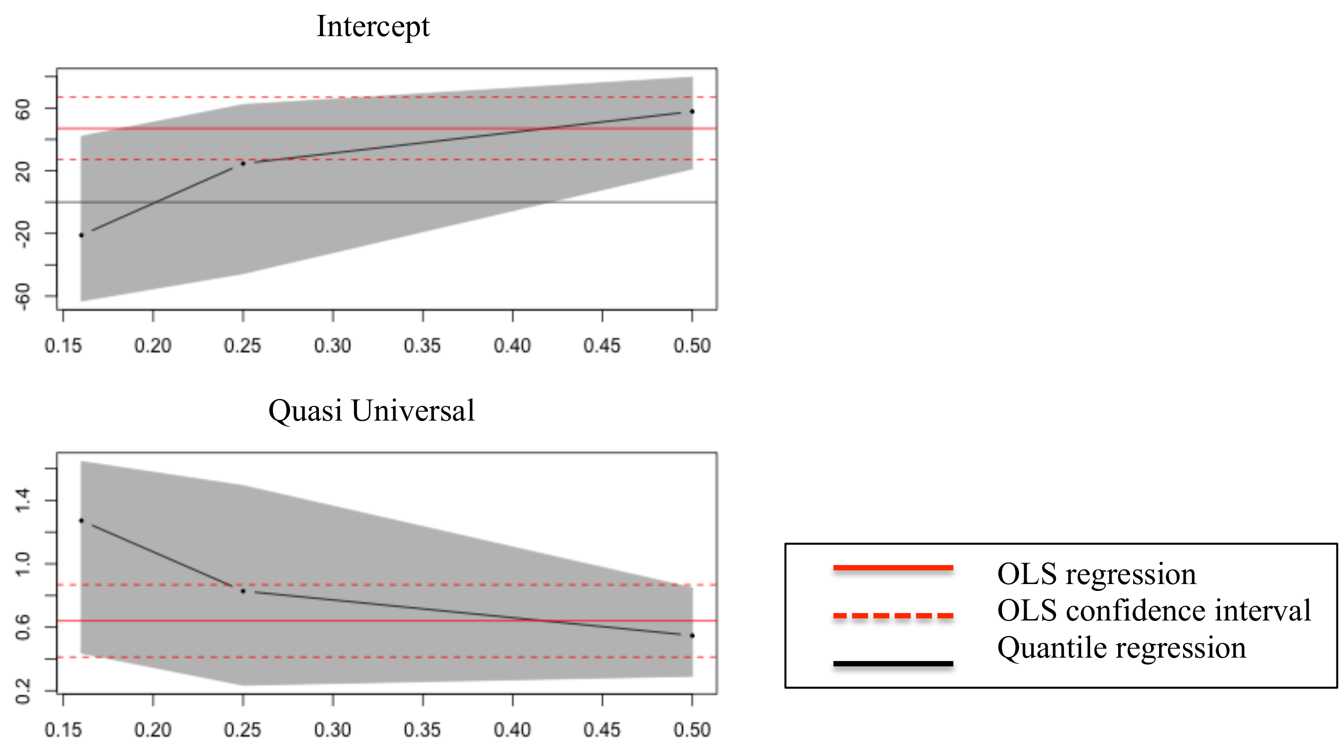

Figure 3: Quantile Regression for Quasi Universal Task

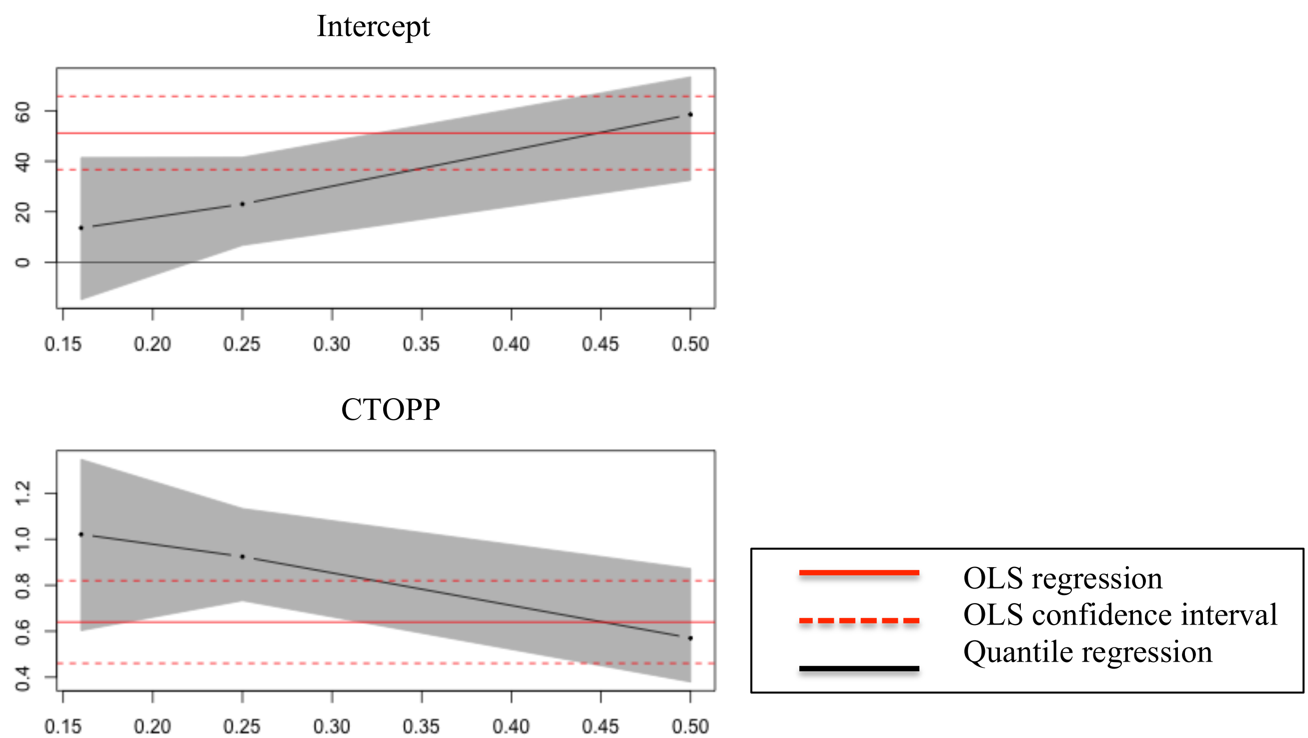

Figure 4: Quantile Regression for CTOPP Task

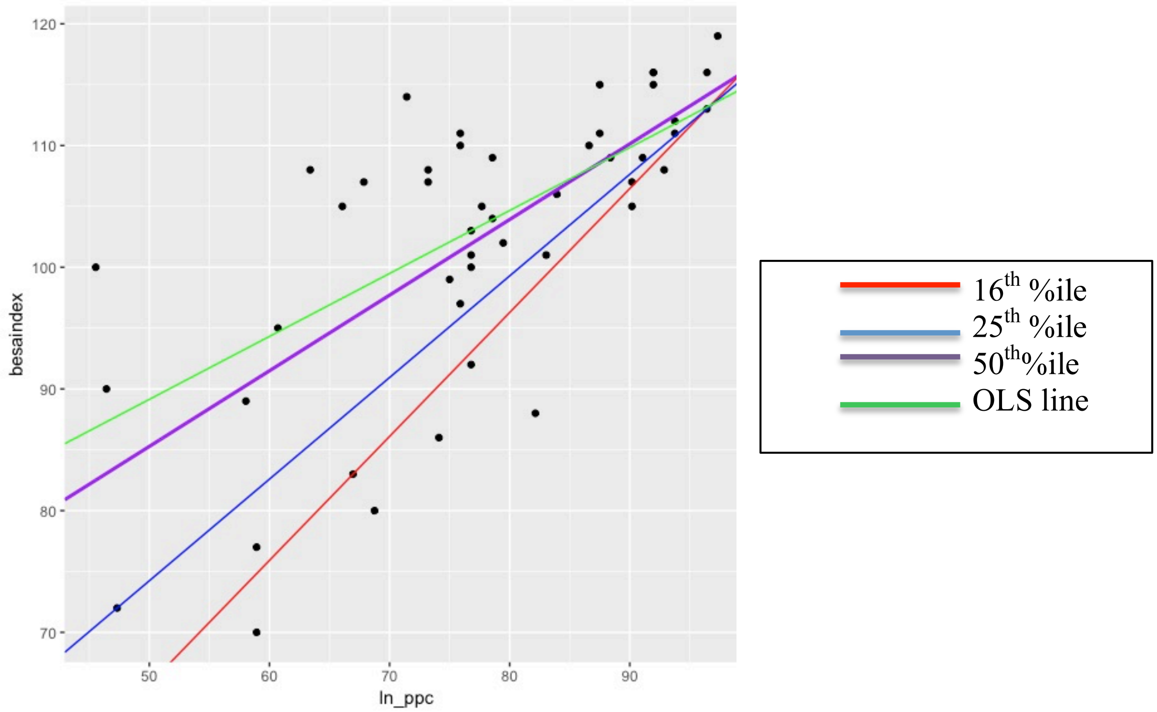

Figure 5: OLS v Quantile Regression: Language Neutral

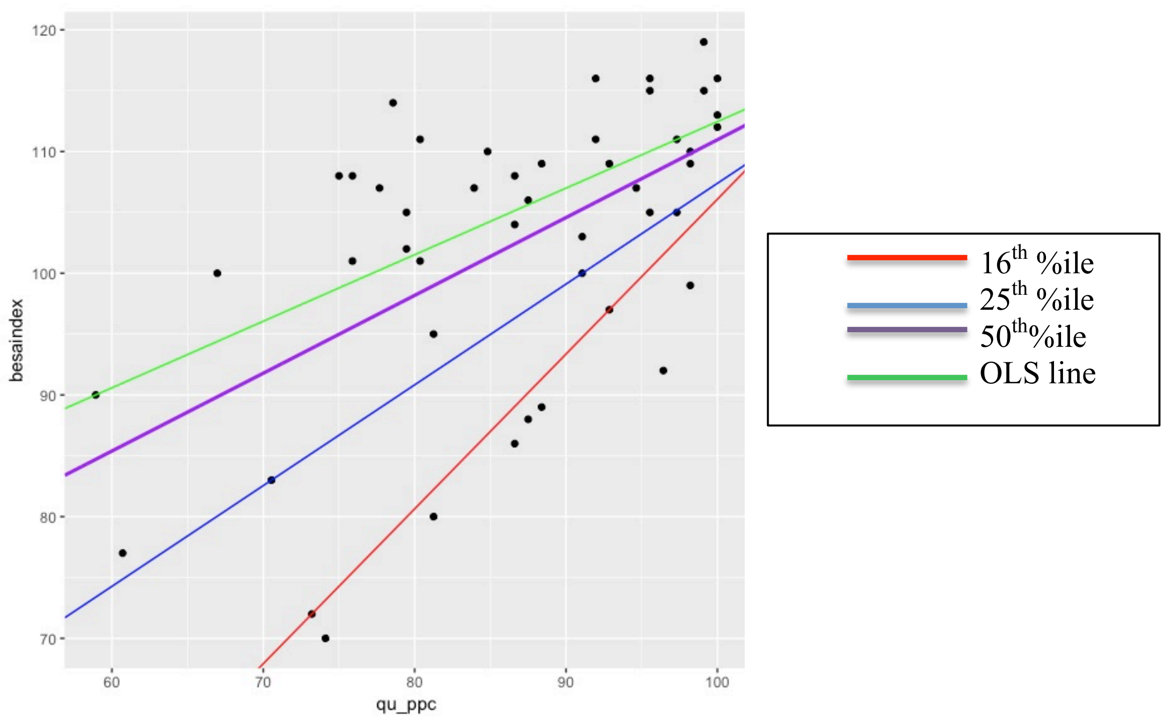

Figure 6: OLS v Quantile Regression: Quasi Universal

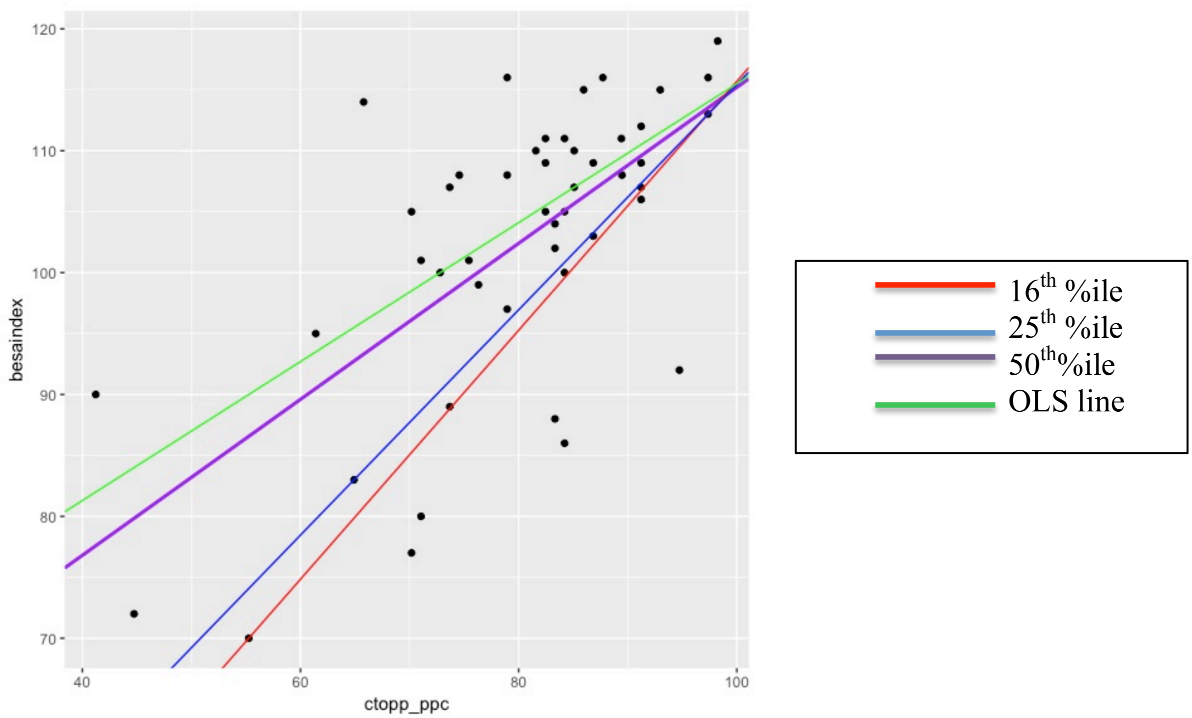

Figure 7: OLS v Quantile Regression: CTOPP

Tables

Table 1: Phonemic Inventory of Spanish and English

	Unique to English	Common to Both	Unique to Spanish
Consonants:	ŋ, dʒ, v, θ, ð, z, ʒ, h, ɹ, j	m, n, w*, p*, b, t*, d, k*, g, tʃ, f, s, l , ɾ, r, ʃ**	ɲ, ʝ, ʎ**, x
Vowels:	æ, ɒ, ɔ, ɛ, ə, ɪ, ʌ, ʊ	e, ɑ, i, o, u, aɪ, aʊ, eɪ, ɔɪ, oʊ	eu, je, ja, jo, ju, wi, we, wa, wo

*phonemic boundaries/manner of production differ substantially
**only found in some dialects

Table 2: Developmental Norms for Speech Sounds

	Early (<4;0)	Middle (4-6)	Late (>6;0)
Spanish:	p, b, m, f, t, d, n, tʃ, ɲ, k, x	r, ɾ, s, g	
English:	m, n, h, w, p, b, t, d, k, g, f-	j, -f, v, θ (F), ð (F) tʃ (F), dʒ (F), l-, -l (F)	ŋ, θ (M), ð (M), s, z, ʃ, tʃ (M), dʒ (M),-l (M), ɹ

F=female; M=male
English norms from Smit, Hand, Freilinger, Bernthal, & Bird (1990)
Spanish norms from Goldstein (2007)

Table 3: Diagnostic Accuracy of Prior Literature

Study	LR+	LR-	Sensitivity	Specificity
Boerma et al. (2015)	Monolingual: 8.3 (QU) 13.3 (Language Specific) Bilingual: 11.9 (QU) 9.0 (Language Specific)	Monolingual: .19 (QU) .08 (Language Specific) Bilingual: .14 (QU) .4 (Language Specific)	Monolingual: 83% (QU) 93% (Language Specific) Bilingual: 83% (QU) 63% (Language Specific)	Monolingual: 90% (QU) 93% (Language Specific) Bilingual: 93% (QU) 93% (Language Specific)
Girbau & Schwartz (2008)	LR+ 9.00	LR- .2	82%	91%
Guiberson & Rodríguez (2013)	2.74	.39	71%	74%
Gutiérrez-Clellen & Simon-Cereijido (2010)	9.71	.62	English task: 55% Spanish task: 61%	English task: 82% Spanish task: 82%
Kohnert Windsor, & Yim (2006)	5.07	LR- .08	66%	92%
Paradis, Schneider, & Duncan (2013)	Combination NWR + ALDeQ + TEGI + ENNI 11.38	Combination NWR + ALDeQ + TEGI + ENNI 0.09	91%	92%
Thordardottir & Brandeker (2013)	Bilingual 4.04	Bilingual 0.18	Bilingual 85% Mono 92%	Bilingual 79% Mono 100%
Van der Linden (2008)	8.57	.21	57% J	33%
Windsor, Kohnert, Lobitz & Pham (2010)	English NWR four syllable: 2.2 Spanish NWR: 3.14	English NWR four syllable: .09 Spanish NWR five-syllable: .52	English NWR Mono 76% Bilingual 94% Spanish NWR Mono 77% Bilingual 58%	English NWR Mono 73% Bilingual 57% Spanish NWR Mono 64% Bilingual 82%

Table 4: Participant and Task Information for Prior Research

Study	LI n (age)	TD n (age)	NWR Used	Scoring Used
Boerma et al. (2015)	Mono: 30 (6;0) Bi: 30 (6;6)	Mono: 30 (6;0) Bi: 29 (7;10)	Dutch-based task (Rispens & Baker, 2012) QU task	Percent Item Correct Percent Phonemes correct Dollaghan & Campbell scoring
Girbau & Schwartz (2008)	11 (8;10)	11 (9;1)	Newly-developed NRT 20 nonwords; 4 @ each syllable level, 1-5	Word level scoring Analysis of errors
Guiberson & Rodríguez (2013)	21 (3;11)	23 (4;1)	NWR task from Ebert et al. (2008):	Dollaghan & Campbell Scoring
Gutiérrez-Clellen & Simon-Cereijido (2010)	49 (5;11)	95 (6;1)	Spanish NWR task 2-4 syllable prosodic and phonological characteristics of Spanish Dollaghan & Campbell's (1998) task	Dollaghan & Campbell Scoring Spanish influenced errors were not counted as wrong
Kohnert Windsor, & Yim (2006)	Mono English 28 (10;6)	Mono English 50 (10;7) Bilingual Spanish-English 22 (9;9)	Dollaghan & Campbell task	Dollaghan & Campbell scoring
Paradis, Schneider, & Duncan (2013)	26 (5;9)	152 (5;10)	CTOPP	Standard scoring- whole item
Thordardottir & Brandeker (2013)	14 Bilingual (5;0) Mono (5;2)	14 Bilingual (4;9) Mono (5;0)	French NWR test (Courcy, 2000)	Dollaghan and Campbell scoring
Van der Linden (2008)	14 (6;5)	15 (6;4)	Dutch-based NWR task developed by deBree (2007)	Whole item and PCC scoring Additions counted as correct Voicing substitutions counted as correct
Windsor, Kohnert, Lobitz & Pham (2010)	Mono English 34 (8;9) Bi 19 (8;5)	Mono English 69 (8;7) Bi 65 (8;0)	Dollaghan & Campbell NRT Spanish task from Ebert, Kalanek, Cordero & Kohnert (2008)	Dollaghan & Campbell scoring

Table 5: Study Design of Prior Research

Study	Design	Independent testing	Blinded testing
Boerma et al. (2015)	2-gate	No	No
Girbau & Schwartz (2008)	2-gate	No	No
Guiberson & Rodríguez (2013)	2-gate	N/A—only Spanish	No
Gutiérrez-Clellen & Simon-Cereijido (2010)	2-gate	No, but tested on different days, order counterbalanced	No
Kohnert Windsor, &Yim (2006)	2-gate	No	No
Paradis, Schneider, & Duncan (2013)	2-gate	No	No
Thordardottir & Brandeker (2013)	2-gate	Yes	No
Van der Linden (2008)	2-gate	N/A—only NWR	No
Windsor, Kohnert, Lobitz, & Pham (2010)	1-gate	No	No

Table 6: Participant Descriptive Statistics

	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation Statistic
Child's age (in months)	60	83	71.19	6.91
SES (parents' years of education)	6	23	12.00	3.65
Percentage English	20.04	79.59	49.35	18.03
Child's Nonverbal IQ	86	140	106.13	14.20

Table 7: Participant Age of First Exposure

		Frequency	Percent
Valid	0*	17	36.2
	1	2	4.3
	2	5	10.6
	3	8	17.0
	4	13	27.7
	5	2	4.3
	Total	47	100.0

*exposed to both languages from birth

Table 8: Assessment Sessions

Prior to Assessment	**Session 1: English (30-45 minutes)**	**Session 2/3: Spanish (35-50 minutes)**	**Session 2/3: English (35-50 minutes)**
● Parents sign consent form ● Parents fill out BESA Questionnaire to determine language history and current language use (BIOS) ● Parents fill out child history questionnaire	● *Reynolds Intellectual Assessment Scales: What's Missing?* (10-15min.) *Odd-Item Out* (10-15 min.) ● Hearing Screening (5-10 min.) ● NWR Task 1 (QU, LN, or CTOPP) randomized (5 minutes)	● BESA-Spanish (30-45 minutes) Phonology Morphosyntax Semantics ● NWR Task 2 (QU or LN) randomized (5 minutes)	● BESA-English (30-45 minutes) Phonology Morphosyntax Semantics ● NWR Task 2 (QU, LN, or CTOPP) randomized (5 minutes)

Table 9: CTOPP Task Items

Item #	**Number of Syllables**	Sound not found in Spanish	Late developing sounds (in English; >6;0)
1	1	✓	✓✓
2	1	✓	✓✓
3	1		
4	1	✓✓	✓
5	1	✓	✓
6	1		
7	1		
8	1		
9	1	✓	
10	2	✓	✓
11	2		
12	2	✓✓	✓
13	3	✓	✓✓
14	3	✓✓✓	✓✓✓✓✓
15	3	✓✓	✓
16	4	✓✓✓✓	✓
17	4	✓✓	✓✓
18	4	✓✓✓	✓✓✓
19	5	✓✓✓✓	✓✓

Table 10: Quasi Universal Task Items

Item #	**Number of Syllables**	Sound not found in Spanish	Late developing sounds (in English; >6;0)
1	2	✓	✓
2*	2		✓
3	2		
4	2		✓
5	3		✓✓
6	3		
7	3		✓
8**	3		✓
9**	4	✓	✓✓
10**	4		✓
11	4		✓✓
12	4		✓✓
13	5		✓✓
14	5		✓✓
15	5	✓	✓✓
16	5		✓

*contains real word in English
** contains real word in Spanish

Table 11: Language Neutral Task Items

Item	**Number of Syllables**	Sound not found in Spanish	Late developing sounds (in English; >6;0)
nafu	2		
muke	2		
bofu	2		
hagi	2		
gapowu	3		
mupoha	3		
hepogu	3		
takedu	3		
wupohake	4		
bohanedu	4		
podunake	4		
kafomuha	4		
tenagiwomu	5		
dubawoteko	5		
fekuhanebo	5		
kewunamogi	5		

Table 12: Descriptive Statistics

	Minimum Statistic	Maximum Statistic	Mean Statistic	Std. Deviation
Spanish Phonology	1	13	8.81	3.67
Spanish Cloze	.1	13	6.09	3.92
Spanish Sentence Repetition	.1	13	6.11	3.39
Spanish Morphology	52	113	80.68	17.34
Spanish Receptive	2	15	9.62	3.25
Spanish Expressive	1	15	9.00	4.00
Spanish Semantic	60	120	96.47	16.44
English Phonology	.1	14	9.28	4.45
English Cloze	3	13	9.36	3.27
English Sentence Repetition	.1	13	8.81	3.69
English Morphology	60	115	96.19	16.62
English Receptive	2	14	10.60	2.36
English Expressive	.1	16	10.36	3.53
English Semantic	62	123	102.47	13.41
BESA English Index	61	119	99.26	14.47
BESA Spanish Index	57	111	88.21	15.83
BESA Index	70	119	102.36	12.05

Table 13: Comparison of NWR Task Performance

Syllable Length	Tasks	p-value
2	CTOPP v. QU CTOPP v. LN	< .001 < .001
3	QU v. LN QU v. CTOPP	< .001 < .001
4	LN v. QU CTOPP v. QU LN v. CTOPP	< .001 < .001 < .001
5	QU v. LN QU v. CTOPP	< .001 = .02

Table 14: Comparison of OLS and Quantile Regression

Variable	**OLS Intercept (SE)**	**Quantile Intercept (SE)**			**p-value**		
		.16	.25	.5	.16 vs. .25	.16 vs. .5	.25 vs. .5
Language Neutral	102.36** (.62)	93.72** (1.02)	97.19** (.84)	103.36* (.52)	.24	.01*	.02*
Quasi Universal	102.36** (.64)	88.95** (1.27)	96.24* (.83)	105.10* (.55)	.07	.02*	.31
CTOPP	102.36** (.64)	95.20** (1.02)	96.89** (.92)	104.06* (.57)	.51	.04*	.06

**p <.05*
***P< .001*

Appendix

Instruments Utilized

Reynolds Intellectual Assessment Scales (RIAS); (Reynolds & Kamphaus, 2009) Nonverbal Intelligence Index (NIX). The RIAS Nonverbal Intelligence Index is an individually administered test utilizing two-subtests: *What's Missing* and *Odd-Item Out*. The NIX assesses nonverbal intelligence by measuring reasoning and spatial ability using novel situations. The NIX has a mean of 100 and a SD of 15.

Bilingual English Spanish Assessment (BESA); (Peña, Gutiérrez-Clellen, Iglesias, Goldstein, and Bedore, 2014). The BESA is a norm-referenced assessment of language ability in Spanish and English. It is normed for children aged 4;0 to 6;11 with varying levels of bilingualism. The test was specifically designed to distinguish between children whose low English ability is due to lack of exposure versus those who are showing evidence of language impairment. The BESA assesses a child's ability in semantics, morphology, and phonology. Each section of the BESA is scored with a mean of 100 and a standard deviation of 15. The morpho-syntax and semantics sections are combined into a Language Index score, with a mean of 100 and a standard deviation of 15. In clinical settings, the Language Index score is calculated using the higher score on each of the subtests, regardless of language tested. This ensures that a child's higher performance is assessed. This is important because bilingual children, especially school-aged children, often exhibit mixed dominance, with language skills differing across domains. For example, a school-age child may have stronger English skills in semantics but stronger Spanish skills in morphosyntax. For the statistical analyses of this study, two additional Language Index scores were calculated: one for only the Spanish subtests, and one for the English subtest. Therefore, statistical analyses could include one of three Index scores: the BESA Index, the English BESA Index, or the Spanish BESA Index.

In this study, the BESA Index score was used as an indication of language impairment. The developers of this test determined that diagnostic accuracy of this measure by identifying children with language impairment on the basis of language sample measures, parent and teacher report, and clinical observation. Using these criteria, they determined that the BESA Index (using the child's stronger language in the semantics and morphosyntax subtest) corresponded to a Sensitivity of 88.9% for 5-year olds and 96% for 6-year olds. The specificity was 84.9% for 5-year olds, and 92.4% for 6-year olds. The corresponding likelihood ratios were LR+ of 5.88 and 11.32, respectively, and LR- of 0.13 and 0.15, respectively (Peña et al., 2014). These values correspond to a LR+ in the suggestive to good range and a LR- in the good range (Dollaghan & Horner, 2011).

Quantile Regression

Quantile regression examines the relationship between the predictor-outcome variables at different quantiles along the outcome variable. While OLS regression examines the mean of Y conditional on X, a quantile regression examines the relation of X to Y, conditional on percentiles of the mean of Y. In other words, it examines whether the strength of the relationship between the variables (in this case, between the NWR tasks and the BESA Language Index) depends on the quantile at which the NWR was examined (how the relationship of X and Y changes depending on the score of Y). Quantile regression is an appropriate method to use when examining data that may not meet the assumptions of normality. Thus, the regression procedure is more robust against outliers. Due to the more complex relationship between the variables (NWR, phonological memory, language experience and age of first exposure), it is likely that a simple regression line would miss more subtle relationships between the variables. The quantile regression function uses all data points to estimate the relationship at each quantile by applying a weight matrix to the data. Therefore, data points that are closest to a certain quantile are weighted more heavily.

Bilingual Children's Language Assessment Study

Bilingual Input-Output Survey

We are interested in your child's history of language exposure for each year of their life. **For each year**, circle the appropriate selection.

What language did you and your family use **at home**:

From	**0-1 year old?**	Spanish	English	Both
	1-2 years old?	Spanish	English	Both
	2-3 years old?	Spanish	English	Both
	3-4 years old?	Spanish	English	Both
	4-5 years old?	Spanish	English	Both
	5-6 years old?	Spanish	English	Both
	6-7 years old?	Spanish	English	Both

At what age was your child first exposed to English outside of the home on a regular basis?

____

Please circle all that apply:

Age:	Attend (please circle):	Language heard/spoken (please circle):	Comments:
0-1	Daycare/Preschool/Other	Spanish/English/Both	
1-2	Daycare/Preschool/Other	Spanish/English/Both	
2-3	Daycare/Preschool/Other	Spanish/English/Both	
3-4	Daycare/Preschool/Other	Spanish/English/Both	
4-5	Daycare/Preschool/Other	Spanish/English/Both	
5-6	Daycare/Preschool/Other	Spanish/English/Both	
6-7	Daycare/Preschool/Other	Spanish/English/Both	

Bilingual Children's Language Assessment
Home Language Profile

DURING THE WEEK

We are interested in what a typical day during the week is like for your child. What activities s/he participates in, who s/he interacts with and what language(s) s/he uses and hears.

For each hour **that your child is awake**, circle the number corresponding to Spanish, English, or both:

	Communication Partner	**Language(s)**					
	(parent, sibling, peer)	Communication Partner – **Input** (what language is being spoken with your child)			Child-**Output** (what language is your child speaking)		
Time		**Spanish**	**Both**	**English**	**Spanish**	**Both**	**English**
6am		☐	☐	☐	☐	☐	☐
7am		☐	☐	☐	☐	☐	☐
8am		☐	☐	☐	☐	☐	☐
9am		☐	☐	☐	☐	☐	☐
10am		☐	☐	☐	☐	☐	☐
11am		☐	☐	☐	☐	☐	☐
12pm		☐	☐	☐	☐	☐	☐
1pm		☐	☐	☐	☐	☐	☐
2pm		☐	☐	☐	☐	☐	☐
3pm		☐	☐	☐	☐	☐	☐
4pm		☐	☐	☐	☐	☐	☐
5pm		☐	☐	☐	☐	☐	☐
6pm		☐	☐	☐	☐	☐	☐
7pm		☐	☐	☐	☐	☐	☐
8pm		☐	☐	☐	☐	☐	☐
9pm		☐	☐	☐	☐	☐	☐
10pm		☐	☐	☐	☐	☐	☐
11pm		☐	☐	☐	☐	☐	☐

DURING THE WEEKEND

We are interested in what a typical day during the weekend is like for your child. What activities s/he participates in, who s/he interacts with and what language(s) s/he uses and hears.

For each hour **that your child is awake**, circle the number corresponding to Spanish, English, or both:

	Communication Partner	**Language(s)**					
	(parent, sibling, peer)	Communication Partner – **Input** (what language is being spoken with your child)			Child-**Output** (what language is your child speaking)		
Time		**Spanish**	**Both**	**English**	**Spanish**	**Both**	**English**
6am		☐	☐	☐	☐	☐	☐
7am		☐	☐	☐	☐	☐	☐
8am		☐	☐	☐	☐	☐	☐
9am		☐	☐	☐	☐	☐	☐
10am		☐	☐	☐	☐	☐	☐
11am		☐	☐	☐	☐	☐	☐
12pm		☐	☐	☐	☐	☐	☐
1pm		☐	☐	☐	☐	☐	☐
2pm		☐	☐	☐	☐	☐	☐
3pm		☐	☐	☐	☐	☐	☐
4pm		☐	☐	☐	☐	☐	☐
5pm		☐	☐	☐	☐	☐	☐
6pm		☐	☐	☐	☐	☐	☐
7pm		☐	☐	☐	☐	☐	☐
8pm		☐	☐	☐	☐	☐	☐
9pm		☐	☐	☐	☐	☐	☐
10pm		☐	☐	☐	☐	☐	☐
11pm		☐	☐	☐	☐	☐	☐

Scoring for Parent Questionnaire—Percentage English

During the Week:

1. Count the number of hours that the child is awake (the hours filled out by parents: _________ (HRSAWAKE)
2. Multiply HRSAWAKE by 10*: _________ (HRSAWAKEWEEK)
3. Input:
    a. Count the number of checkmarks for English: _________ (HRSEH)
    b. Count the number of checkmarks for BOTH: _________ (HRSBTH)
    c. Multiply HRSBTH by 0.5: _________ (HRSBOTHWEEK)
    d. Add HRSEH and HRSBOTHWEEK: _________ (HRSINPUTWEEK)
4. Output:
    a. Count the number of checkmarks for English: _________ (HRSEH)
    b. Count the number of checkmarks for BOTH: _________ (HRSBTH)
    c. Multiply HRSBTH by 0.5: _________ (HRSBOTHWEEK)
    d. Add HRSEH and HRSBOTHWEEK: _________ (HRSOUTPUTWEEK)
5. Add HRSINPUTWEEK and HRSOUTPUTWEEK: _________ (HRSEHWEEK)
6. Multiply by 5: ________ (HRSTOTALWEEK)

During the Weekend:

1. Count the number of hours that the child is awake (the hours filled out by parents: _________ (HRSAWAKE)
2. Multiply HRSAWAKE by 4: _________ (HRSAWAKEENND)
3. Input:
    a. Count the number of checkmarks for English: _________ (HRSEH)
    b. Count the number of checkmarks for BOTH: _________ (HRSBTH)
    c. Multiply HRSBTH by 0.5: _________ (HRSBOTHEND)
    d. Add HRSEH and HRSBOTHWEEK: _________ (HRSINPUTEND)
4. Output:
    a. Count the number of checkmarks for English: _________ (HRSEH)
    b. Count the number of checkmarks for BOTH: _________ (HRSBTH)
    c. Multiply HRSBTH by 0.5: _________ (HRSBOTHEND)
    d. Add HRSEH and HRSBOTHWEEK: _________ (HRSOUTPUTEND)
5. Add HRSINPUTWEEK and HRSOUTPUTWEEK: _________ (HRSEHEND)
6. Multiple by 2: __________(HRSTOTALEND)

Add HRSTOTALWEEK and HRSTOTALEND and divide by (HRSAWAKEWEEK + HRSAWAKEENND)

Multiple by 100= _________________ Percentage English

* 5 days per week/ 2 days per weekend, multiplied by 2 to represent language input and output