January 2015

Nonword Repetition And Sentence Repetition In Preschool Children: A Comparison Of Speech And Language Screening Measures

Rachel Therrien

NONWORD REPETITION AND SENTENCE REPETITION IN PRESCHOOL CHILDREN: A COMPARISON OF SPEECH AND LANGUAGE SCREENING MEASURES

by

Rachel R. E. Therrien
Bachelor of Arts, Brandon University, 2011

A Thesis
Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of requirements

for the degree of

Master of Science

Grand Forks, North Dakota
May
2015

This thesis, submitted by Rachel Therrien in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and hereby approved.

Alycia Cummings, CCC-SLP, Ph.D.

John P. Madden, CCC-A, Ph.D.

Sarah Robinson, CCC-SLP, Ph.D.

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Wayne Swisher
Dean of the School of Graduate Studies

April 22, 2015

PERMISSION

Title	Nonword Repetition and Sentence Repetition in Preschool Children: A Comparison of Speech and Language Screening Measures
Department	Communication Sciences and Disorders
Degree	Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in her absence, by the Chairperson of the department of the Dean of the School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Rachel Therrien
April 22, 2015

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	**Meaning**
AGS	*AGS Early Screening Profiles* (American Guidance Service, 1990)
CELF-P2 %	*Clinical Evaluation of Language Fundamentals – Preschool: Second Edition* (Semel, Wiig, & Secord, 2004) percentile ranks
CELF-P2 SS	*Clinical Evaluation of Language Fundamentals – Preschool: Second Edition* (Semel, Wiig, & Secord, 2004) scaled scores
HAPP-3	*Preschool Phonological Screening* from the *Hodson Assessment of Phonological Patterns – Third Edition* (Hodson, 2004)
LI	language impairment
N	number of participants
NRT	nonword repetition task
NRT nonwords	percent of correclty produced nonwords on the nonword repetition task
P	participant
SR	severity rating
SRT	sentence repetition task
SRT sentences	percent of correctly produced sentences on the sentence repetition task

SRT words	percent of correctly produced words on the sentence repetition task
SSD	speech sound disorder

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to express my sincere appreciation to the members of my advisory committee for their guidance and support during my time in the master's program at the University of North Dakota

ABSTRACT

Nonword repetition tasks and sentence repetition tasks have been shown to accurately differentiate children with speech and/or language disorders from their typically developing peers (Dollaghan & Campbell, 1998; Archibald & Gathercole, 2006). Additionally, they provide information on semantic, syntactical, and phonological processing as well as memory and acoustic perceptual skills (Vance, Stackhouse, & Wells, 2005). Despite their utility as a screening measure, their use in clinical practice has been minimal. Instead, practicing speech-language pathologists often choose to use publisher-provided screening measures or their own developed measures, rather than using research-based screening methods that have recognized differential diagnosis ability (Nash, Leavett, & Childs, 2011). The purpose of the present study is to compare the current speech and language screening measures used by student clinicians in the UND Speech, Language, and Hearing Clinic to the Dollaghan and Campbell's (1998) nonword repetition task (NRT) and the sentence repetition task (SRT) taken from the *Clinical Evaluation of Language Fundamentals – Preschool Second Edition (CELF-P2)* (Semel, Wiig, & Secord, 2004). In order to determine the clinical usefulness of the SRT and NRT as a speech and language screener, potential cut-off points to distinguish typically developing (TD) children and those who may have language impairment (LI) and/or speech sound disorder (SSD) were also considered. Forty-nine English-speaking, monolingual, preschool-aged children (ages 3;1 to 5;9 years) were recruited for this study. Results indicated a limited correlation was found between UND's measures and the SRT and the NRT. In addition, a strong relationship was found between the SRT CELF-P2 scaled scores and the percent consonant correct (PCCs) and percent phoneme correct (PPCs) scores of the SRT and NRT. Overall, a greater number of participants performed more poorly on the SRT and NRT than on UND's screening measures. These results indicate that the SRT and NRT may be detecting speech and/or language disorders that UND's measures fail to detect.

CHAPTER I

INTRODUCTION

Developmental communication disorders, such as speech sound disorders (SSD) and language impairment (LI), significantly affect an individual's education, future employment, and overall well-being (ASHA, 2008). Approximately 7% of all children in the United States have LI. Sixty-one percent of school-based speech-language pathologists (SLPs) have indicated they have served individuals with LI (ASHA, 2008). These children have deficits in language acquisition and use despite having normal hearing and nonverbal intelligence. For example, children with LI may have impairments in lexical, grammatical, and morphological development (Graf Estes, Evans, & Else-Quest, 2007).

It is estimated that 5-8% of children in the United States have a functional SSD. Approximately 80% of these children will require speech therapy intervention and 50-70% will exhibit general academic difficulty through high school (Bernthal, Bankson, & Flipsen, 2012; Williams, McLeod, & McCauley, 2010). Given the number of children impacted by LI and SSD and the importance of early intervention, it is necessary to develop tools that can assist in the assessment and diagnostic process.

Identification of children with SSD and LI becomes critical around the age of 4 to 5 years. Intervention prior to beginning school is necessary, as untreated childhood speech and language impairment can have negative consequences throughout a child's life

(Mcleod & Harrison, 2009). A comprehensive assessment of a child's expressive and receptive language abilities as well as their phonological and articulation skills should be completed before a diagnosis is made (Bernthal et al., 2012; Paul & Norbury, 2012). Unfortunately, these assessments often take a lot of time to complete. Furthermore, not every child needs a full evaluation. Therefore, many SLPs often use screening measures as a first step in distinguishing children who have SSD or LI from those who do not. These mass screenings are often done in schools or medical settings and identify children who require a more comprehensive evaluation (Paul & Norbury, 2012). In order to continue to evaluate the accuracy of these screenings to ensure proper identification of LI and SSD, further research must be conducted. Therefore, the purpose of the present study is to compare the current speech and language screening measures used by student clinicians in the University of North Dakota (UND) Speech, Language, and Hearing Clinic to the Dollaghan and Campbell's (1998) nonword repetition task (NRT) and the sentence repetition task (SRT) taken from the *Clinical Evaluation of Language Fundamentals – Preschool Second Edition (CELF-P2)* (Semel, Wiig, & Secord, 2004).

Screening measures for speech. A combination of assessment tools is recommended to provide multiple views of a child's speech (Johnson, Weston, & Bain, 2004). Furthermore, these assessments should be efficient yet thorough to offer a reasonable basis for decision making (Tyler et al., 2002). Typical screening measures are nonstandardized or standardized and range from asking a child to answer a few questions to collecting a full speech sample.

Nonstandardized screening measures are inexpensive but have the drawback of being unable to compare the child's performance to normative data. An example of nonstandardized screening is engaging a child in conversation or having them count to ten. Standardized screening measures used in educational settings usually consist of having the child point to or name a picture or object. They have the advantage of comparing a child's performance to standardized norms but are often costly to purchase and focus on one task (i.e., naming) rather than using a wide variety of other methods of evaluation. For example, the *Hodson Assessment of Phonological Patterns – Third Edition* (HAPP-3) (Hodson, 2004) (Appendix A) screener involves having a child name twelve different picture items (e.g., "glasses"). The clinician then transcribes the child's productions and compares these productions to those of typically developing (TD) children. A clinician relying solely on naming tasks should consider other methods of assessment as well to incorporate a wide variety of cognitive skills, as naming tasks require the involvement of different cognitive skills than other methods of assessment, such as repetition tasks (Vance, Stackhouse, & Wells, 2005).

Cognitive processes involved in naming versus repetition tasks. The cognitive processing required for naming involves having the child rely on his/her own mental representations in order to name pictures. The first step in naming is accessing stored knowledge of the visual representation (e.g., an image of a cat). If the image is recognized, the child will then access their semantic knowledge of the word and stored phonological information. For the example mentioned above, the child must consider the semantic representation of what a "cat" is, and then access his knowledge of the sounds used to create the word "cat" and how these sounds are articulated. The last step in this process is using his motor programming to produce the word "cat."

In contrast, repeating words involves different levels of processing. In word-repetition, there is no picture for the child to name and he must rely solely on the acoustic signal. For example, the child will first hear the word "cat." To repeat this word, the child must first remember the word by storing it in his short-term memory. Then the child must access stored phonological representations of the word form he just heard. Here he may also access his semantic knowledge of the word "cat," or he may only use bottom-up processing and consider the sounds that compose the word. The child will then use his knowledge of how the sounds in the word "cat" are articulated, keeping in mind that he has already received a model provided by the acoustic signal. The child will then use motor programming skills to produce the word "cat" (Vance, Stackhouse, & Wells, 2005).

Screening measures for language. Screening procedures used to differentiate TD children from those with LI are similar to speech screenings in that using a variety of tools is recommended (ASHA, 2008). In addition, nonstandardized and/or standardized tasks may be used. An example of a nonstandardized language screener may require a child to label what he sees in a picture book. In an educational setting, one of the standardized screenings used is known as the *AGS Early Screening Profiles* (American Guidance Service, 1990) (Appendix B). This screener involves having the child point to or label a variety of picture items (e.g., naming an elephant). Similar to the HAPP-3 speech screener, this language screener relies heavily on naming, and therefore it places different cognitive demands on the child than tasks such as repetition tasks.

Nonword repetition as a speech and language screener. Developmental speech and language screening measures can involve a number of approaches. One approach that may be useful in detecting SSD and/or LI is using a nonword repetition task. Two sets of nonword stimuli have been used most frequently in research focused on nonword repetition. The first set consists of 40 nonwords developed by Gathercole and Baddeley (1990) that were later revised to form the *Children's Test of Nonword Repetition* (Gathercole et al. 1994). The second set consisted of 16 nonwords that were developed by Dollaghan and Campbell (1998). These 16 nonwords are commonly referred to as the *Nonword Repetition Test* (NRT) (Appendix C) and have been shown to accurately differentiate children with LI and/or SSD from TD children (Dollaghan & Campbell, 1998; Vance, Stackhouse, & Wells, 2005; Shriberg et al., 2009; Deevy et al., 2010).

Thus far, research has shown utility in the use of NRT for the detection of LI (Dollaghan & Campbell, 1998; Marton & Schwartz, 2003; Archibald and Gathercole, 2006; Gathercole, 2006; Graf Estes, Evans, & Else-Quest, 2007; Archibald & Joanisse, 2009). The majority of these studies have taken into consideration age, cultural background, socioeconomic status, and educational background when distinguishing between children with LI and TD children. Their findings have been consistent in that children with and without LI perform differently on these tasks.

Previous research has demonstrated that groups of children with LI and/or SSD repeat nonsense words less accurately than do their peers who have typically developing speech and language skills (Dollaghan & Campbell, 1998; Sahlen et al., 1999; Archibald & Gathercole, 2006). A lack of repetition accuracy in these children may, in part, be due to cognitive processing differences. For instance, children with LI may perform poorer on NRTs than their age-matched peers due to their limitations in simultaneously processing information. For these children, NRT accuracy decreases when word length and task complexity increases (Marton & Schwartz, 2003). Marton and Schwartz (2003) investigated the role of working memory in language comprehension using NRT. Specifically, they compared children with LI to TD children. They found that working memory plays an important role in language comprehension. The children with LI had a more limited working memory capacity than their age-matched peers. They perfomed more poorly on all working memory tasks involving NRT.

Despite its clinical utility, research has shown that nonword stimuli should not be the only tool used when screening for LI in young children, as it may not take into account the different types of deficits associated with LI (e.g., deficits in grammaticality, listening comprehension, etc.) (Deevy et al., 2010). For example, Bishop et al. (2006) found that, when conducting a study on identical twins at high risk for LI, some sets of twins did poorly on NRT but did well on tense/agreement morpheme use tasks while other sets of twins did well on NRT but showed deficits in tense/agreement morpheme use. These findings suggest that NRT may not be identifying every child who has LI.

Sentence repetition as a speech and language screener. Another task that may be useful in identifying children with LI and/or SSD is a sentence repetition task (SRT). Children who have more difficulty accurately repeating verbally presented sentences are much more likely to have LI than children who can repeat sentences verbatim (Nash, Leavitt, & Childs, 2011). Moreover, children with LI have less accurate SRT scores, particularly when the number of words to be recalled increases. For example, children with LI tend to have difficulties with grammatical morphemes (e.g. plural –s as in "dog**s**") (Eisenberg & Guo, 2013). Furthermore, these children have more difficulties processing sentence structure and tense marking than children who do not have LI. (Montgomery, Evans, & Gillam, 2009).

Due to its ability to differentiate children with LI from TD children, sentence repetition has been used for a number of years as a core subtest in many standardized language batteries (Archibald & Joanisse, 2009). For example, the *Clinical Evaluation of Language Fundamentals – Preschool: Second Edition (CELF-P2)* (Semel, Wiig, & Secord, 2004) contains a recalling sentences subtest where the child must recall thirteen sentences of increasing length (Appendix D). In addition, Conti-Ramsden & Botting (2001) reported that, out of four measures (i.e. nonword repetition, sentence recall[1], third-person singular, and past tense), SRT was found to be the best clinical marker of LI and had a sensitivity[2] rating of 90% and a specificity[3] rating of 85%. Although SR has been used in the field for many years, its use as a language screener has been limited.

Research involving SRT with children with SSD is virtually nonexistent. This may be due to the current focus on LI and repetition tasks. In studies that do focus on SRT and SSD, it is generally assumed that children do not also have LI (Newton, 2012). In addition, research on intelligibility that does include children with SSD has focused on single words rather than on sentences or connected speech (Flipsen, 2006). Phonetically transcribing single words verses entire sentences is easier and takes a shorter amount of time. Therefore, the lack of research dedicated to SRT may be due to the cumbersome analysis process.

---

[1] Sentence recall corresponds to sentence repetition in Cont-Ramsden & Botting's (2001) study.
[2] Sensitivity refers to the ability of a task to correctly identify children with LI.
[3] Specificity refers to the ability of a task to correctly identify children who have typically developing language skills.

When research has compared phoneme production accuracy in longer utterances and single words, the results have been interesting. One study found that some children with speech difficulties perform well on single word articulation tests, but have poor intelligibility in connected speech measures (Grunwell, 1987). In addition, Newton (2012) conducted a study in which three males (ages 11;8 to 12;10), who were currently in speech therapy, were required to complete a sentence repetition task. The results indicated that participants demonstrated articulation and phonology errors within each of the sentences. It was noted that these participants tended to make more errors in connected speech than in single words. Given that SRTs typically include sentences of increasing length, these types of tasks may be able to identify children with SSD who may otherwise pass a speech screening task that uses single word stimuli.

Nonword versus sentence repetition. According to Conti-Ramsden & Botting (2001), sentence repetition involves short-term memory as well as prior language knowledge (i.e. semantics, syntax, morphology). In contrast, repeating nonwords requires a prior knowledge of phonology and eliminates the need to rely on one's prior experience with vocabulary words. In addition, nonword repetition also places some demands on a child's short-term memory, as nonwords of increasing length must be remembered (Archibald & Gathercole, 2006; Deevy et al., 2010). Both tasks require the participant to be able to accurately perceive and produce the stimuli.

Nonword repetition and sentence repetition have been shown to accurately differentiate children with SSD and/or LI from their TD peers. Additionally, they provide information on semantic, syntactical, and phonological processing as well as memory and acoustic perceptual skills. Despite their utility as a screening measure, their use in clinical practice has been minimal. Instead, practicing speech-language pathologists often choose to use publisher-provided screening measures or their own developed measures, rather than using research-based screening methods that have recognized differential diagnosis ability. Thus, the purpose of the present study is to compare the speech and language screening measures used by the UND Speech, Language, and Hearing Clinic (i.e., the *Hodson Assessment of Phonological Patterns – Third Edition (HAPP-3)* (Hodson, 2004) and the AGS Early Screening Profiles (American Guidance Service, 1990)) with the Dollaghan and Campbell's (1998) NRT and the SRT taken from the *Clinical Evaluation of Language Fundamentals – Preschool Second Edition (CELF-P2)* (Semel, Wiig, & Secord, 2004).

To determine whether the SRT and NRT were more accurate in distinguishing TD children from those with LI and/or SSD than the screeners used by UND clinicians, two questions were asked: 1) *Do children perform in a similar way on the SRT and NRT and UND's speech and language screeners?,* and 2) *Are there specific cut-off points that can be identified to distinguish TD children and those who may have LI and/or SSD?*

It was predicted that children would perform better on UND's measures than on the SRT and NRT. Both the HAPP-3 speech screener and AGS language screener contain single word stimuli, and therefore would be easier to produce/recall than the SRT (Grunwell, 1987). Both of UND's measures also contain real words, which have the advantage of accessing stored knowledge of the visual and semantic representations. Nonwords do not have this advantage, which may lead to poorer production performance on the NRT (Vance, Stackhouse, & Wells, 2005).

To answer the second question related to the potential screener cut-off criteria, all analyses completed for the SRT and NRT were taken into consideration. These analyses were then used to determine whether a pattern existed between those children who performed poorly on the SRT and NRT. Determining pass/fail criteria would allow future clinicians to use the SRT and NRT as a screener without needing to complete the laborious and time-consuming analyses incorporated into this study.

CHAPTER II

METHOD

The present study compared the SRT (Semel, Wiig, & Secord, 2004) and NRT (Dollaghan & Campbell, 1998) to UND's speech and language screeners. The specific goals of this study were to: 1) Determine whether children performed similarly on the SRT and NRT and UND's measures, and 2) Determine whether cut-off points could be identified for the SRT and NRT, thus establishing pass/fail criteria.

Participants. Participants were 49 preschool-aged subjects (22 females and 27 males, aged 3;1 to 5;9 years) recruited and tested in conjuction with yearly preschool screenings at preschools in the Grand Forks, ND vicinity. All children were monolingual English speakers. Parents received information regarding the screening and a participation consent form two weeks prior to the screening. Those parents who wished their child attend the screening were asked to send back the consent form to the preschool.

Procedure. Four speech and language screening measures were administered to all participating children: the Nonword Repetition Task (NRT) (Dollaghan & Campbell, 1998), the Sentence Repetition subtest of the CELF-P2 (Semel, Wiig, & Secord, 2004), the *Preschool Phonological Screening* from the HAPP-3 (Hodson, 2004), and the Verbal Concept subtest and Overall Basic Skills subtest from the *AGS Early Screening Profiles* (American Guidance Service, 1990). In addition, a hearing screening was performed on all participants.

Nonword repetition task (NRT). Each participant sat in front of a computer which aurally presented a recording of 16 nonsense words created by Dollaghan and Campbell (1998) (Appendix C). These nonwords were arranged in a hierarchy beginning with single syllables and ending with four-syllable nonwords The participant was asked to listen to “the alien” (i.e., the computerized voice) and then repeat what he or she heard after each word. Approximately 2 seconds were given for the child to repeat each word. A word was repeated one time if the child did not did not respond to the first presentation.

Sentence repetition task (SRT). The participant sat in front of a computer on which 13 short sentences taken from the CELF-P2 were aurally presented. The sentences ranged from 3 to 13 words in length (Appendix D). The participant was asked to listen to each sentence and then repeat what he or she heard. Approximately 2-5 seconds was given after each sentence with additional time provided if a sentence needed to be repeated. These sentences were pre-recorded by the clinician to ensure uniformity and pilot tested on 34 undergraduate students to determine the degree of naturalness of each sentence. The 13 sentences were recorded, in a random order, three times by the clinician for a total of 37 sentences (2 sentences were taken out due to poor audio quality). The 34 undergraduate students were then asked to rate each sentence on a scale from 1 to 5. A rating of one indicated that the sentence sounded very unnatural and a rating of 5 indicated that the sentence sounded very natural. The mean rating for each sentence was calculated and the sentences with the highest average rating were used as the SRT stimuli.

UND screening measures. In addition to the NRT and SRT, the children also completed the standard speech, language, and hearing screening administered by the graduate student clinicians from the UND clinic. Each child’s voice and fluency were informally assessed throughout the screening process. The student clinicians screened three communication areas:

1) *Speech*. The speech screening used by the clinicians was the *Preschool Phonological Screening* from the HAPP-3 (Hodson, 2004). This measure involved showing a participant 10 pictures and having the participant label each picture (Appendix A). Specific speech sounds in each word were targeted (e.g., in "fork" - the child's production of "f", "r", and "k" is assessed). The total possible number of speech errors a participant may have produced was 23.
2) *Language*. The language screening measures used by the clinicians were taken from the *AGS Early Screening Profiles* (American Guidance Service, 1990). This measure was divided into different subsets. The Overall Verbal Concept had a maximum total raw score of 25. It comprised the Verbal Receptive Concepts subtest, which had a raw score of 11, and the Verbal Expressive Concepts, which had a raw score of 14. The Overall Basic Skills also had a total raw score of 25. It consisted of the Basic Skills Receptive subtest, which had a raw score of 14, and the Basic Skills Expressive, which had a raw score of 11 (Appendix B).

   For these measures, the participant was asked to point to or verbally label picture items. For example, “show me the elephant” was a receptive task in the Verbal Concept subtest and the child had to point to a picture of an elephant out of a field of four pictures. An example of an expressive task for this subtest was to have the child label a picture, such as “book”. For the Basic Skills subtest, an example of a receptive task was to have the child “point to the biggest stick” when shown a picture of sticks of various sizes. An expressive task for this subtest was, for example, showing the child a picture of 5 squirrels and having them count the squirrels out loud.

3) *Hearing*. The hearing screening followed standard audiometric screening procedures (American National Standards Institute, 1991). Standard hearing screening involves presenting single tones of three different frequencies (1000 Hz, 2000 Hz, and 4000 Hz) at 20 dB to both the right and left ears of each child. For the purpose of this study, children were allowed to participate in the study if they did not pass the screening at one frequency in one ear (either the left or right ear).

Data analysis. After the NRT was administered and narrowly transcribed, three measures were completed. The first measure was calculating a percent phoneme correct (NRT PPC) to determine the number of phonemes (i.e., consonants and vowels) a child produced accurately in each of the 16 nonwords. In this simple binary measure, each phoneme in a word was scored as either correct or incorrect (Dollaghan, Biber, & Campbell, 1993). For example, if the nonword was /dɔɪf/ and the child produced [dɔɪt], two out of the three phonemes were scored as correct. Any omitted or added phonemes were also included in the total number of phonemes and scored as incorrect. For example, if the nonword was /dɔɪf/ and the child produced [dɔɪfi], that child would have correctly produced three out of four phonemes. Therefore, the total number of phonemes varied across participants as certain children added and/or omitted phonemes while others did not. The number of correct phonemes in each of the 16 nonwords were then summed and divided by the total number of phonemes (i.e., 96) to receive a percentage.

The percent consonant correct (NRT PCC) score was calculated in a similar manner to the NRT PPC score, though in this case only the consonants were scored. Thus, the 56 consonants produced in the 16 nonwords were scored as either correct or incorrect (Shriberg, 1997). Similar to the PPC score, consonant omissions and additions were scored as incorrect. For example, if the nonword was /teɪvɑk/ and the child's production was [eɪvɑt], one out of three consonants would be scored as correct.

Finally, each nonword was scored in a binary fashion, as either being fully correct or incorrect and a percent of correctly produced nonwords was calculated. A nonword was considered incorrect if one or more of the phonemes in a nonword was produced incorrectly. For example, if the word /tɑʊʤ/ was produced as [tɑʊd], that nonword would be scored as incorrect. Similarly, if the nonword /ʧoʊvæg/ was produced as [toʊbi], it would also be scored as incorrect. This measure was used to determine whether this simple binary measurement was comparable to the results of the PCC and/or PPC scores. That is, it was used to determine whether those participants who received a low percentage on the PCC and the PPC scores also received a low percentage on the total number of nonwords produced correctly and vice versa.

Sentence repetition task (SRT). After the SRT task was completed and the sentences were phonetically narrowly transcribed, five measures were used to analyze the data. The first two measures were used to calculate a sentence level percent phoneme correct (SRT PPC) and a sentence level percent consonant correct (SRT PCC), which was calculated in the same manner as the NRT PPC and PCC scores. The total number of target phonemes for the SRT PPC score was 369 and the total number of consonants for the SRT PCC score was 232. Again, the total may have changed depending on the productions of the participant. For example, if the sentence /ðeɪ pleɪ wɪθ blɑks/ ("They play with blocks.") was produced as [pleɪ wɪft blɑks] ("Play wift blocks."), the child would receive a score of 10 out of 14 phonemes correct for the SRT PPC score and 7 out of 10 consonants correct for the SRT PCC score. In this example, the child omitted the word "they" and substituted the "f" in "with" for "th". Adult-like pronunciations of function words, such as "and" were considered correct. For example, if a child produced "and" as [æn] ("an") or [ɛn] ("n"), it would be scored as correct. In addition, all phonemes in a word were scored as incorrect if the word was not in the correct order in a sentence. For instance, if the sentence "The play castle was built by the girls and boys." was produced as "The play castle was built by the boys and girls.", the phonemes in "boys" and "girls" were scored as incorrect.

The third measure consisted of calculating an overall word percentage accuracy score based on the 111 words of the SRT. A word was considered incorrect if it contained a minimum of one grammatical error. For instance, if the sentence "The play castle was built by the girls and boys." was produced as "The play castle was builded by the girl and boy.", a total of 7 out of 10 words were scored as correct. Here, the child made a past tense error (i.e., "builded") and omitted the plural –s in "boys" and "girls." Speech sound production errors were not calculated as incorrect. However, errors of this nature could have influenced the overall accuracy score. For example, if a child was unable to produce a /s/ or /z/ sound, plural –s would have been directly affected and considered incorrect if it was omitted.

In addition, each of the 13 sentences was scored in a binary fashion as being either correct or incorrect as a whole, and a percent of correctly produced sentences was calculated. For example, the above sentence would have been scored as incorrect, whereas the sentence "They play with blocks' contains no errors and, therefore, would have been scored as correct.

Finally, the SRT was scored in accordance to the CELF-P2 manual. For this task, a raw score was obtained by counting the number of errors produced by the child in each sentence (a higher score equals less errors). Following the CELF-P2 scoring procedures, a word was considered to be in error if it was omitted or if it contained a grammatical error. For instance, if the sentence "Didn't the boys eat the apples?" was produced as "Didn't boy eat the apples?", the sentence would be considered to have two errors: one word omission (i.e., "the") and one omission of plural –s (i.e., in "boys"). For the first two sentences, a score of 2 was given if the sentence contained no errors, a score of 1 was given if the sentence contained one error, and a score of 0 was given if the sentence contained two or more errors. For sentences three through thirteen, a score of 3 was given if a sentence contained no errors, a score of 2 was given if a sentence contained one error, a score of 1 was given if a sentence contained two to three errors, and a score of zero was given if a sentence contained four or more errors. A raw score was calculated and was then converted to a CELF-P2 subtest scaled score and a percentile rank, which allowed for a comparison to standardized norms distinguishing TD children from children with LI.

Reliability transcription. Every pre- and post-treatment assessment session was digitally recorded using the internal sound field microphones of an Olympus Digital

Voice Recorder (DS-2) or RCA Digital Voice Recorder. Each audio sample was then digitally transferred to a PC laptop for transcription using a speech recording computer program, Adobe Audition®. Highly trained and blinded research assistants used the International Phonetic Alphabet (IPA) to narrowly transcribe all speech samples. Based on these transcriptions, the data were organized for phonological analyses according to target sound/cluster and word position. Approximately 25% of each child's NRT and SRT speech sample were reliability-checked by a second transcriber; these speech samples were compared for point-by-point consonant agreement that was required to reach an agreement threshold of 85% for each sample. If this threshold was not reached, the speech sample in question was re-transcribed until two transcribers reached the designated threshold. Due to time constraints, only the NRT was analyzed. Overall, the transcriber reliability for the NRT was 81% (range: 65-92%). As this did not meet threshold requirements, transcription will be further analyzed until an agreement threshold is met.

Comparison measurements. After all screening procedures were analyzed separately, the SRT and NRT were compared with UND's speech and language screening measures.

The first step in comparing this study's measures with UND's measures was to determine whether a relationship existed between UND's speech and language screeners and the PPC and PCC scores of the NRT and SRT. For the NRT, a variety of correlational analyses were completed separately for the NRT PPC and PCC measures. The following analyses were performed:

- HAPP-3 score and the NRT
- NRT and the AGS verbal concepts receptive subtest

- NRT and the AGS verbal concepts expressive subtest
- NRT and the AGS verbal concepts total raw score
- NRT and the AGS basic skills receptive subtest
- NRT and the AGS basic skills expressive subtest
- NRT and the AGS basic skills total raw score

A similar set of correlational analyses was completed for the SRT PPC and PCC measures:

- HAPP-3 score and the SRT
- SRT and the AGS verbal concepts receptive subtest
- SRT and the AGS verbal concepts expressive subtest
- SRT and the AGS verbal concepts total raw score
- SRT and the AGS basic skills receptive subtest
- SRT and the AGS basic skills expressive subtest
- SRT and the AGS basic skills total raw score

The second comparison took place to determine whether there was a relationship between the CELF-P2 SRT scaled scores (SS) and the PCC and PPC scores of the SRT and NRT as well as the overall (OV) accuracy on the SRT and the NRT. In addition, the SRT SS and the UND measures were compared. To determine if these relationships existed, correlational analyses were completed for the following tasks:

- SRT SS and HAPP-3
- SRT SS and AGS verbal concepts total raw score
- SRT SS and AGS basic skills total raw score
- SRT SS and NRT PCC

- SRT SS and NRT PPC
- SRT SS and SRT PCC
- SRT SS and SRT PPC
- SRT SS and OV SRT
- SRT SS and OV NRT

This study also attempted to determine cut-off points for pass/fail criteria for the NRT and/or SRT. Three different measures were used in an attempt to identify cut-off scores. The CELF-P2 SRT subtest scaled score was used to determine whether a pattern existed for children who fell below the average range for scaled scores (average range is from 7 to 13) and percentile ranks (average range is from 16 to 84). In addition, a PCC severity rating was assigned for the SRT PCC and the NRT PCC. According to Shriberg & Kwiatkowski (1982), a PCC of 85-100% indicates mild impairment for speech sound production, 65-85% mild-moderate, 50-65% moderate-severe, and <50% severe impairment. Furthermore, children who produced less than twenty-five percent of the words and/or sentences correctly on the SRT and the NRT were also examined to determine whether the 25% cut-off point could be used.

CHAPTER III

RESULTS

In order to determine pass/fail criteria, results of the UND speech and language screenings and the SRT and NRT were analyzed independently. These tasks were then compared to one another. Participants who performed poorly on the screenings were also analyzed in further detail. In agreement with the predictions of this study, the analyses revealed differences between UND's measures and the SRT and NRT. In addition, possible cut-off points were examined.

UND Standard Screening. The results of the UND speech, language, and hearing screening are summarized in Table 1 below. A child was considered to have passed the AGS Early Screening Profiles if he or she achieved a score of three or above on the Language Screening Index (LSI). To determine an LSI, the child's age and raw scores for the receptive and expressive components of the Verbal Concepts subtest and Basic Skills subtests were used and converted into a Point Score. This score was then converted into an LSI. All participants passed the language portion of the screen.

In order to pass the HAPP-3 speech screener, factors such as the child's age, number of articulation errors, and use of phonological processes (e.g. fronting – substituting [t] for /k/ or [d] for /g/) were taken into consideration. For example, a child who was unable to correctly produce or omitted velar consonants (e.g. /k/ and /g/) and liquid consonants (e.g. /l/ and /r/), would have failed the screening. In this sample, 47 out of 49 participants passed the speech screener.

For hearing, 40 out of 49 participants passed the standard audiometric hearing screening. As mentioned previously, children were allowed to participate if they passed all but one frequency in one ear. Voice and fluency were also informally assessed during the screening process; no voice or fluency difficulties were noted for any of the 49 participants during the screening process.

Table 1. Results of standard speech, language, and hearing screenings.

	PASS	**FAIL**
HAPP-3	N = 47	N = 2
AGS Early Screening Profiles	N = 49	N = 0
Hearing	N = 40	N = 9

Sentence repetition (SRT) and nonword repetition (NRT). For the SRT, the total number of correctly produced words (N = 111 words) were calculated and a percent accuracy obtained. None of the 49 participants received a percentage of <25% for correctly repeating single words in sentences. The same calculations were completed for the SRT sentences (N = 13 sentences). For the 13 sentences, 22 out of 49 participants received a percentage of <25% accuracy. The total number of correctly produced nonwords (N = 16) for the NRT was also calculated and a percentage obtained. Out of the 16 nonwords, 10 out of 49 participants obtained a percentage of <25% accuracy. Results are summarized in Table 2 below. Interestingly, the eight participants who were below normal limits for the CELF-P2 published norms also obtained an overall percentage accuracy of less than 25% correct for the SRT sentences (i.e. ≤ 3 correct sentences) and less than 50% correct for the NRT (i.e. ≤8 correct nonwords). In addition, out of the ten participants who received a percentage of <25% for correctly repeating nonwords, eight participants also received a percentage of <25% for correctly repeating sentences; the remaining two participants received a percentage of <50% for the sentences.

Table 2. Range of percent correct scores for the CELF-P2 sentences and words in the SRT and nonwords in the NRT.

Percentage Accuracy Scored	CELF-P2 Sentences	CELF-P2 Words	NRT Words
75% - 100%	N = 1	N = 25	N = 0
50% - 74%	N = 11	N = 21	N = 17
25% - 49%	N = 15	N = 3	N = 22
<25%	N = 22	N = 0	N = 10

PCC and PPC scores were calculated for both the SRT and the NRT. The mean, standard deviation (SD), and highest (max.) and lowest (min.) percentage obtained by the participants for the NRT and SRT PCC and PPC scores are presented in Table 3. In general, PCC and PPC scores were similar for each participant across tasks (Table 4). In addition, a strong correlation was found between the NRT and SRT PCC scores and the overall number of correct nonwords or sentences (Table 4). This indicates that a participants speech sound production accuracy (i.e., their PCC score) was directly related to their repetition accuracy on the SRT and NRT. Furthermore, PCCs and PPCs were typically lower for participants who scored lower in other SRT and NRT measures.

Table 3. Descriptive statistics for percent consonant correct (PCC) and percent phoneme correct (PPC) in NRT and SRT.

	SRT PCC	SRT PPC	NRT PCC	NRT PPC
Mean	73.30	73.77	67.39	73.22
SD	15.45	15.31	15.67	13.58
Max.	96.98	97.02	94.64	95.83
Min.	28.75	27.42	23.21	19.39

Table 4. Correlation coefficients (r) comparing SRT and NRT PPC and PCC scores as well as overall (OV) SRT and NRT scores with PCC scores.

	Correlation Coefficients
SRT PPC and PCC scores	r = .99 ($p < .001$)
NRT PPC and PCC scores	r = .93 ($p < .001$)
SRT and NRT PPC scores	r = .68 ($p < .001$)
SRT and NRT PCC scores	r = .64 ($p < .001$)
OV SRT and SRT PCC scores	r = .78 ($p <.001$)
OV NRT and NRT PCC scores	r = .81 ($p <.001$)

In addition to these descriptive statistics, a PCC severity score (Shriberg & Kwiatkowski, 1982) was assigned to each participant's SRT and NRT performance. The results of the PCC Severity Scale for this study are summarized below in Table 5. For the SRT-PCC, nine participants were moderate-severe and four were severe. For the NRT-PCC, nine participants were moderate-severe and seven were assigned a severe rating.

Table 5. Number of participants that fell under each category according to the percent consonant correct (PCC) severity scale. A PCC of 85-100% indicates mild impairment for speech sound production, 65-85% mild-moderate, 50-65% moderate-severe, and <50% is considered severe impairment (Shriberg & Kwiatkowski, 1982).

	SRT	NRT
Mild (85% - 100%)	N = 14	N = 4
Mild-Moderate (65% - 85%)	N = 22	N = 29
Moderate-Severe (50% - 65%)	N = 9	N = 9
Severe (<50%)	N = 4	N = 7

The SRT, a subtest from the CELF-P2, was also scored in accordance with the CELF-P2 manual. Raw scores were converted into scaled scores and percentile ranks. According to standardized norms, typically developing children have scaled scores in a range of 7 to 13 for scaled scores, which convert to percentile ranks between 16 and 84. There were eight participants involved in this study that scored below the normal limits, indicating their language skills were below those of typically developing children (Table

6). This is in contrast to the AGS Early Screening Profiles used to detect LI, as all 49 participants passed the AGS screener indicating that they all had typically developing language skills.

Table 6. Summary of eight participants scoring below the average range for scaled scores and percentile ranks on the CELF-P2 SRT.

Participant	HAPP-3	AGS	CELF-P2 SS	CELF-P2 %
P3	Pass	Pass	5	5
P8	Fail	Pass	4	2
P9	Pass	Pass	6	9
P13	Pass	Pass	5	5
P15	Pass	Pass	5	5
P27	Pass	Pass	5	5
P40	Pass	Pass	4	2
P46	Pass	Pass	5	5

In addition to the above analyses, correlational analyses were completed using the SRT and NRT PCC and PPC scores and all screening measures used by the UND student clinicians. The results of these analyses are summarized in Table 7 below. A correlation coefficient (r) of <.1 indicates no correlation between variables, .1 to .29 indicates a small correlation, .3 to .49 indicates a medium correlation, and ≥.5 indicates a strong correlation (Gravetter & Wallnau, 2013).

A medium positive correlation was found between the SRT PCC and PPC scores and the AGS verbal concepts subtest total raw score. This indicates there was a relationship between participants' speech sound production accuracy on the SRT and their naming accuracy on the verbal concepts subtest. Children who received a higher PCC and PPC score on the SRT also received a higher score on the verbal concept subtest and vice versa.

In addition, a medium positive correlation was found between the overall SRT score (i.e., the number of sentences produced correctly) and the AGS verbal concept subtest. A relationship was found between participants' ability to accurately repeat sentences and their ability to name picture vocabulary. This indicates that there is a connection between a child's vocabulary knowledge and their ability to retain and accurately repeat a contextually-based sentence.

A medium positive correlation was also found between the SRT and the NRT PPC and PCC scores and the AGS basic skills subtest total raw score. This also reveals a relationship between speech sound production accuracy on the SRT and naming accuracy on the basic skills subtest. Children performed similarly on the SRT and NRT in terms of PPC and PCC scores, as they did on the basic skills subtest. Therefore, there may be a connection between a child's speech sound production accuracy on a repetition task, and their naming accuracy on a naming task such as the tasks used in the AGS screening measures.

A weak negative correlation was found between the SRT PCC and PPC scores and the NRT PPC scores and the HAPP-3 as well as the overall SRT score and the HAPP-3. A weak positive correlation was also found between the overall NRT score and the HAPP-3 scores and the NRT PCC scores and the HAPP-3 scores (Table 6). This indicates that participants who performed well on the HAPP-3 performed poorly on the SRT and NRT PCC and PPC scores. Given that the SRT and NRT overall scores and the PCC and PPC scores are strongly correlated with other measures, this may indicate that the HAPP-3 is not identifying all children with a SSD.

Table 7. Comparison of HAPP-3 and AGS screening measures with SRT and NRT PCC and PPC scores and overall (OV) SRT and NRT scores.

	SRT PCC	**SRT PPC**	**NRT PCC**	**NRT PPC**	**OV SRT**	**OV NRT**
HAPP-3	$r = -.13$	$r = -.12$	$r = .03$	$r = -.06$	$r = -.13$	$r = .14$
Verbal Concepts Total	$r = .48$ ($p < .001$)	$r = .48$ ($p < .001$)	$r = .27$	$r = .23$	$r = .45$ ($p < .001$)	$r = .08$
Basic Skills Total	$r = .30$ ($p < .05$)	$r = .32$ ($p < .05$)	$r = .32$ ($p < .05$)	$r = .32$ ($p < .05$)	$r = .14$	$r = .20$

A correlational analysis was also completed between the CELF-P2 SRT SS and the PCC and PPC scores of the SRT and the NRT, the overall SRT and NRT scores, and UND's screening measures. The results are summarized in Table 8 below. A strong correlation was found between the CELF-P2 SRT SS and the SRT PPC and PCC scores. A strong correlation was also found between the overall SRT score and the CELF-P2 SS. As these measures involve the same task, there is no variation between tasks. Therefore, the relationship is solely dependent on the different analyses used. A strong correlation between these analyses indicates that participants are performing in a similar manner when the task is analyzed in terms of language (i.e., scaled scores and overall accuracy score) or speech sound production accuracy (i.e., PPC and PCC scores). In addition, this suggests that the transcription and scoring methods of the SRT PPC and PCC scores and the overall SRT score are fairly reliable and valid.

For the NRT, a strong correlation was found between the CELF-P2 SRT SS and the PPC and PCC scores of the NRT as well as the CELF-P2 SRT SS and the overall NRT score. This indicates that the SRT and NRT may both be capable of measuring language and/or speech sound deficits. Participants who received low scores on the SRT scaled scores, which measures LI, also received low scores on the NRT PPC and PCC scores, which measures LI and SSD. In addition, overall production accuracy of the nonwords on the NRT was related to the scaled scores on the SRT. Therefore, the overall NRT score and the NRT PPC and PCC scores may be useful in detecting LI and SSD as they correlate to the CELF-P2 SRT SS language measurement.

A medium positive correlation was found between the AGS verbal concepts subtest total and the CELF-P2 SRT SS. Participants scored similarly on the verbal concepts naming task and the CELF-P2 SRT SS repetition task. However, all participants passed UND's language screening, whereas eight participants fell below average for the CELF-P2 SRT SS. As the AGS verbal concepts subtest total is combined with the basic concepts subtest total to establish the passing criterion on the AGS screener, this may indicate that children are passing UND's language screener because of their basic concept scores.

Table 8. Comparison of the CELF-P2 scaled scores (SS) with SRT and NRT PCC and PPC scores, overall (OV) SRT and NRT scores, and HAPP-3 and AGS screening measures.

	CELF-P2 SRT SS
Verbal Concepts Total	r = .30 ($p < .02$)
Basic Skills Total	r = .18
HAPP-3	r = -.06
SRT PCC	r = .77 ($p < .001$)
SRT PPC	r = .75 ($p < .001$)
NRT PCC	r = .66 ($p < .001$)
NRT PPC	r = .62 ($p < .001$)
OV SRT	r = .71 ($p <.001$)
OV NRT	r = .51 ($p <.001$)

In order to continue to establish pass/fail criteria, the eight participants who fell below normal limits according the CELF-P2 SRT published norms and the two participants (i.e., P2 and P8) who failed the HAPP-3 speech screener (nine participants in total) were examined in closer detail to see whether possible error response patterns emerged (Table

9). Participant eight (i.e., P8) failed the HAPP-3 speech screener and fell below the CELF-P2 average norms, which indicates he was identified as potentially having both speech and language deficits. Out of the eight participants who fell below average on the CELF-P2 SRT scoring measures, the majority received a severity rating of moderate-severe or severe for both the SRT and NRT PCC scores. In addition, all eight participants produced less than 50% of the total number of SRT sentences and NRT nonwords correctly.

The second participant (i.e., P2) who failed the HAPP-3 speech screener produced fewer than 25% of the SRT sentences correctly. However, he received a PCC severity rating of mild-moderate on the SRT and mild on the NRT, indicating he was not as severe in his speech sound production severity rating as the eight participants who fell below average for the SRT CELF-P2 scoring measures. This disparity in speech sound production ratings may be due to the difference in speech sound targets in each of the tasks. For example, the HAPP-3 contains liquids (i.e., /r/ and /l/) whereas the NRT does not target these sounds.

Table 9. SRT and NRT PCC scores with corresponding PCC severity ratings of the two participants who failed the HAPP-3 screener, and the eight participants who fell below average according to CELF-P2 scoring measures (i.e., percentile ranks and scaled scores). The percent of correctly produced words/sentences on the SRT and NRT are also included.

Participant	SRT words	SRT sentences	NRT nonwords	SRT PCC	SR	NRT PCC	SR
P2*	78.38	23.08	68.75	80.08	Mild-Mod.	87.50	Mild
P3	45.05	15.38	37.50	45.34	Severe	55.17	Mod.-Sev.
P8**	60.36	7.69	31.25	58.64	Mod.-Sev.	55.36	Mod.-Sev.
P9	67.57	23.08	25.00	68.22	Mild-Mod.	62.07	Mod.-Sev.

Table 9. cont.

Participant	SRT words	SRT sentences	NRT nonwords	SRT PCC	SR	NRT PCC	SR
P13	55.86	7.69	12.50	53.28	Mod.-Sev.	36.21	Severe
P15	59.46	27.27	18.75	56.23	Mod.-Sev.	46.43	Severe
P27	63.96	15.38	25.00	60.33	Mod.-Sev.	39.34	Severe
P40	52.25	7.69	18.75	54.89	Mod.-Sev.	43.33	Severe
P46	61.26	7.69	6.25	61.80	Mod.-Sev.	26.79	Severe

Note: *speech screen fail only; **speech screen and CELF-P2 SRT SS fail

The ten participants who correctly produced fewer than 25% of the nonwords on the NRT and the three participants who correctly produced fewer than 50% of the SRT words were also examined (Table 10). Out of the ten participants who fell below 25% accuracy on their nonword productions, eight out of ten participants also received less than 25% accuracy on their SRT sentence productions (P15 and P38 received a score of 27.27% accuracy and 53.85% accuracy respectively). These participants ranged from mild to severe in their SRT and NRT PCC severity scores.

The three participants who produced fewer than 50% of the SRT words correclty had similar overall accuracy scores on the SRT sentences and NRT nonwords. All three participants had less than 50% accuracy on the SRT words and sentences as well as the NRT nonwords. Their PCC severity scores ranged from mild-moderate to severe. One participant (i.e., P3) also fell below average on the CELF-P2 SRT SS.

Table 10. SRT and NRT PCC scores with corresponding PCC severity ratings of the ten participants who fell below 25% accuracy on the NRT and three participants who fell below 50% accuracy when producing words on the SRT. The percent of correctly produced words and/or sentences on the SRT are also included.

Participant	SRT words	SRT sentences	NRT nonwords	SRT PCC	SR	NRT PCC	SR
P3*	45.05	15.38	37.50	45.34	Severe	55.17	Mod.-Sev.
P5	67.57	15.38	18.75	57.20	Mod. – Sev.	62.53	Mod.-Sev.
P13	55.86	7.69	12.50	53.28	Mod.-Sev.	36.21	Severe
P15	59.46	27.27	18.75	56.23	Mod.-Sev.	46.43	Severe
P22*	27.93	0.00	6.25	28.75	Severe	23.21	Severe
P33	64.86	7.69	18.75	62.61	Mod.-Sev.	67.86	Mild-Mod.
P37*	46.85	0.00	31.25	40.52	Severe	66.07	Mild-Mod.
P38	86.49	53.85	18.75	85.9	Mild	58.93	Mod.-Sev.
P40	52.25	7.69	18.75	54.89	Mod.-Sev.	43.33	Severe
P45	82.88	30.77	12.50	51.04	Severe	41.07	Severe
P46	61.26	7.69	6.25	61.80	Mod.-Sev.	26.79	Severe
P49	72.07	23.08	18.75	69.07	Mild-Mod.	55.17	Mod.-Sev.

Note: *Less than 50% accuracy when producing SRT words; **CELF-P2 SRT SS fail; ***Less than 50% accuracy on SRT words and CELF-P2 SRT SS fail

For the SRT, the 13 sentences were examined individually to determine whether the eight participants who fell below the average range for the CELF-P2 SRT scoring produced specific sentences in error. None of the sentences were produced in error by all of the eight participants. In general, participants did better on the first two sentences than the last two sentences; however this was not true for all participants. In addition, there were no consistent semantic or grammatical errors found for these eight participants. For example, certain participants omitted articles in some sentences, others omitted adjectives (e.g. "brown"), and other participants produced both types of omissions.

The NRT was also examined to see if there was a pattern between the seven participants who were rated as “severe” according to the PCC severity scale (Table 4), and their productions of each of the nonwords. Of these seven participants, four participants (P13, P15, P40, P46) also fell below average on the CELF-P2 SRT SS and scored less than 25% accuracy on the NRT nonwords and the SRT sentences. In general, participants who were rated as “severe” had more difficulties with two-syllable nonwords and were more likely to produce all of the two-syllable nonwords incorrectly. Interestingly, the nonword /nɔɪtaʊf/ was produced incorrectly by all participants rated as “severe”, but produced correctly by the majority of the other participants, including those rated as “moderate-severe”. Most participants had difficulties producing the three- and four-syllable nonwords. Therefore, the three- and four-syllable nonwords were not useful in distinguishing TD children from those who potentially have LI and/or SSD, as all children had difficulties with these nonwords.

Overall, participants performed poorer on the SRT and NRT than on UND’s measures as was predicted. When comparing measures, a relationship was found between the SRT and NRT PPC and PCC scores and the AGS screening measures, indicating that participants performed similarly when the task was naming versus repetition. Results also revealed performance similarities when comparing the CELF-P2 SRT SS and the SRT and NRT PPC and PCC scores. Moreover, correctly producing fewer than 25% of the NRT nonwords and/or the sentences on the SRT might be a possible criterion score to detect LI and/or SSD.

CHAPTER IV

DISCUSSION

A good screening measure is one that is brief, inexpensive, and comprehensive. The nonword repetition task (NRT) (Dollaghan and Campbell, 1998) and sentence repetition task (SRT) (Semel, Wiig, & Secord, 2004) meet these criteria. Both tasks can be administered in less than 10 minutes and the cost to administer them is minimal. Furthermore, when the tasks are combined, they provide information on multiple communication aspects including, but not limited to, syntax, morphology, phonology, articulation, voice, and fluency (Vance, Stackhouse, & Wells, 2005; Marton & Schwartz, 2003; Nash, Leavitt, & Childs, 2011).

This study compared the HAPP-3 speech screener and the AGS language screener to the SRT and NRT. The following questions were asked in order to compare the measures and determine the clinical utility of the SRT and NRT.

1) Do children perform in a similar way on the SRT and NRT and UND's speech and language screeners?
2) Are there specific cut-off points that can be identified to distinguish children who are TD and those who may have LI and/or SSD?

In relation to the first question, overall the participants did better on UND's speech and language screeners than on the SRT and NRT. All participants passed the AGS language screener. This is in contrast to the eight participants who fell below average according to the CELF-P2 SRT scoring measures. The larger number of participants performing poorly on the SRT screener compared to the AGS language screener may be due to the differing nature of the tasks. The AGS language screener is a picture-naming task whereas the SRT is a repetition task. These tasks involve different cognitive processes, which may have influenced the results. Naming requires that the child rely on his own mental representations in order to name pictures. Repeating words involves different levels of processing as there is no picture for the child to name and he must rely solely on the acoustic signal (Vance, Stackhouse, & Wells, 2005).

However, the SRT may in fact be detecting children with LI that the AGS language screener is failing to detect. Eight participants fell below the average range for percentile ranks and scaled scores, which indicates that they, at a minimum, had difficulties processing language information. Many of these children also received low PPC and PCC scores and did poorly on the NRT, meaning they also may be at risk for SSD. It is important to accurately identify children with LI at a young age in order to identify and plan for academic weaknesses that they might experience in their education, which could affect future employment and overall well-being (ASHA, 2008).

Only two participants failed the UND HAPP-3 speech screener. In contrast, seven participants for the NRT were rated as "severe" on the PCC severity scale, and ten participants produced less than 25% of the nonwords correctly. In addition, a limited relationship was found between the HAPP-3 speech screener and the NRT PCC and PPC scores. It is unclear whether the absence of a relationship was due to the inherent differences in these tasks mentioned previously, or the differences these tasks may have in detecting speech sound disorders. For instance, participants should be performing more poorly on the HAPP-3 than the NRT as later-developing sounds (e.g., /r/ as in “rock”) are absent from the NRT and therefore may be easier to produce (Shriberg, 1993). The NRT also contains no sound clusters, whereas three words containing sound clusters, which are more difficult to produce than single sounds, are found in the HAPP-3 screener (e.g. “star”). Alternatively, the NRT contains four-syllable words while the HAPP-3 contains mainly one-syllable words. As single syllable words are easier to produce than multisyllabic words, this may be a further causal factor as to why a limited relationship exists between the two screeners (Dollaghan & Campbell). Despite the varying nature of the tasks, the PCC severity scale has been shown to be an accurate diagnostic classification system for SSD (Shriberg & Kwiatkowski, 1982). That, along with the overall nonword accuracy percentage, signifies that the NRT screener detected a greater number of children who potentially have speech difficulties as compared to the HAPP-3 speech screener.

In general, the NRT and SRT PCC and PPC scores were similar for each participant between tasks, indicating consistency between these measures. Participants who received a low PCC and PPC score on the NRT also received a low PCC and PPC score on the SRT and vice versa. For those participants who did have different PCC and PPC scores on the NRT and SRT, this may be due to several factors that coincide with the limited relationship found between the HAPP-3 speech screener and the SRT PCC and PPC scores. First, the SRT requires that the participant remember and generate sentences of varying length, which places greater cognitive demands than remembering single nonwords. This may impact performance accuracy on the SRT versus the HAPP-3 and NRT as children perform better on single words compared to sentences (Grunwell, 1987).

Unlike the NRT, the SRT contains several sound clusters (e.g. "**sc**hool" and "**pl**ay") as well as later developing sounds such as consonant "r" (e.g. "**r**abbit"). However, the limited relationship between the HAPP-3 and SRT is not explained by the sound inventory as the HAPP-3 also contains later developing sounds. Therefore, the length of utterance (i.e., sentences versus single words) might explain the difference.

A further explanation that may explain differences found between the HAPP-3 SRT and NRT PCC and PPC scores is that the nonwords of the NRT contain no contextual information. The nonwords have no meaning and are not dependent on context. Therefore, the NRT nonwords may be easier to produce, as the participant would not need to rely on his background knowledge of words and their relationship to one another. In addition, participants do not need any prior morphologic or syntactic knowledge making the NRT nonwords and HAPP-3 words easier to produce that the SRT sentences. Overall, it appears that the HAPP-3 has fewer challenging aspects in terms of speech sound production that may explain enhanced performance over other tasks.

A relationship was found between the AGS language screener and the SRT and NRT PCC and PPC scores. Typically, PCC and PPC scores are related to the detection of SSD (Shriberg & Kwiatkowski, 1982). However, these measures are also important in LI detection. There is a well-known association between language and phonological impairments (e.g. Dollaghan & Campbell, 1998). An impairment of phonological storage often accompanies LI (Gathercole, 2006). Thus, children with LI often have difficulties with repetition tasks not only because of the language skills required, but also because of a deficit in their phonological working memory. That is, these children have difficulties properly storing the correct order of phonemes. This has a direct impact on PCC and PPC scores as the accuracy decreases when phonological errors are present. Therefore, a possible explanation for the relationship found between the AGS and the SRT and NRT PCC and PPC scores may be due to the relationship between language and phonological working memory.

The SRT and NRT PCC and PPC scores were strongly correlated with the CELF-P2 SRT SS. For the CELF-P2 SRT SS and the SRT PPC and PCC scores, all measures involved the same task, and therefore this finding was somewhat expected. The relationship was dependent on the different analyses used and not on the differences between tasks. Participants performed similarly when the task was analyzed in terms of language (i.e., scaled scores and overall accuracy scores) and speech sound production accuracy (i.e., PPC and PCC scores).

The relationship found between the CELF-P2 SRT SS and the NRT PPC and PCC scores might indicate that the SRT and NRT are both capable of measuring language and/or speech sound deficits. Participants who received low scores on the SRT scaled scores, which measures language deficits, also received low scores on the NRT PPC and PCC scores, which measures speech sound production deficits. This also coincides with the relationship between the SRT and NRT PCC and PPC scores, as speech sound production accuracy was similar across both tasks. Therefore, the NRT PPC and PCC scores might be useful in detecting LI and SSD, as they correlate to the CELF-P2 SRT SS language measurement as well as the SRT PCC and PPC scores speech sound production measurement.

In contrast, there was a very limited relationship found between the CELF-P2 SRT SS and the screening measures used by the UND Speech, Language, and Hearing Clinic.

There are three possible explanations for why this occurred. The first reason may be that the tasks, as mentioned previously, are inherently different; one task may be easier or more difficult for a child than another task. For example, due to the differing cognitive processes mentioned previously, it may be easier for a child to name a picture than to remember a sentence. Sentence repetition requires the use of phonological working memory, syntactic and morphology skills, and vocabulary knowledge (Gathercole, 2006). This task also has the disadvantage of not being associated with a visual stimulus (i.e., a picture of the object to name).

The second reason may be that UND students did not have the skills to properly administer the tasks used to measure speech and language. Speech-language pathologists often rely on their clinical experience to differentiate TD children from children with LI and/or SSD (Chen, 2011; Francois et al., 2015). Graduate students have limited exposure in the assessment and identification of these disorders. They may not have the tools to accurately identify children with LI and/or SSD. In particular, graduate students vary in their ability to accurately identify speech errors (Cronin et al., 2014). In addition, limited supervision is provided when these screenings take place. Therefore, the limited relationship between the CELF-P2 SRT SS and UND's measures may be, in part, due to variations in student clinician experience.

The third possibility may be that the UND measures are less sensitive at detecting speech and/or language deficits as there was, in general, a limited relationship between their measures and the measures used in this study. A number of participants were identified as having speech and/or language deficits when using both standardized and nonstandardized measurements on the SRT and NRT. This is in stark contrast to the two participants who were identified as having speech errors and the absence of participants who were identified as having language deficits. Based on these analyses it is difficult to conclude that UND's measures are accurately identifying children at risk for LI and/or SSD.

This study attempted to establish a pass/fail criterion for the SRT and NRT measures used as a screener. The eight participants who fell below normal limits, as measured by the CELF-P2 standardized scores, were examined to determine whether a pattern existed between their standardized scores on the SRT and the PPC and PCC measures of SRT and the NRT performance. These same eight children also received a severity rating of "moderate-severe" or "severe" on the SRT and NRT using the PCC severity scale. Furthermore, these children correctly produced fewer than 50% of their SRT sentences and NRT nonwords.

The ten participants who correctly produced fewer than 25% of the NRT nonwords and the three participants who correctly produced fewer than 50% accuracy of the SRT words were also examined to determine a possible pass/fail criterion. Out of the ten participants who fell below 25% accuracy on their nonword productions, eight participants also performed with less than 25% accuracy on their SRT sentence productions. The three participants who correctly produced fewer than 50% of the SRT words also had overall accuracy scores of less than 50% on the SRT sentences and the NRT nonwords. This indicates that the 25% accuracy cut-off on the SRT sentences and NRT nonwords might be a possible pass/fail criterion for these tasks. The overall accuracy of the SRT sentences and NRT nonwords strongly correlates with the SRT and NRT PCC scores as well as the CELF-P2 SRT SS. Therefore, a percent accuracy cut-off would be a simple, fast criterion measure to use as it is comparable to standardized speech and language measures.

Caveats and Future Studies. In order to establish the SRT and the NRT as a speech and language screener, it is important to determine pass/fail cut-off guidelines for clinicians. In this study, producing fewer than 25% of the SRT sentences and NRT nonwords accurately was identified as a possible cut-off guideline. However, future studies should replicate the guidelines found in this study to determine consistency.

In addition, the sample of children used to test the SRT and NRT was limited in terms of size, cultural and educational background, and socioeconomic status. The majority of the children lived in the surrounding area of the preschool and had Caucasian backgrounds. Future studies should consider using a larger sample size with a more diverse background.

Finally, to assess the ability of the NRT and SRT to distinguish between TD children and those with a SSD and/or LI, future studies should include a full assessment of each child including the use of standardized speech and language tests. This will ensure that those children who fail the screener are those who have SSD and/or LI and are not TD children. If all children who fail the screener also receive a diagnosis of LI and/or SSD, this will indicate that the screener is sensitive in detecting these disorders. Moreover, if all children who pass the screener are assessed as being TD, this will indicate that the screener succeeds at eliminating the need for TD children to undergo unnecessary further assessment. A speech and language screener that has sufficient sensitivity and specificity will ensure that children with a wide range of speech and/or language deficits are identified. Accurate identification will provide these children with the proper programming to achieve success in their academic and personal lives.

APPENDICES

Appendix A
HAPP-3 Stimuli

1. boats – /boʊts/
2. fork – /fɔrk/
3. glasses – /glæsɪz/
4. gum – /gʌm/
5. leaf – /lif/
6. rock – /rɔk/
7. soap – /soʊp/
8. spoon – /spun/
9. star – /stɑr/
10. watch – /wɑʧ/
11. zip – /zɪp/
12. nose – /noʊz/

Appendix B
AGS Screener Stimuli

Verbal Concepts Subtest

1. Show bird
2. Show elephant
3. Say spoon
4. Show pencil
5. Say book
6. Say watching TV
7. Say eating
8. Say running
9. Say umbrella
10. Show cart
11. Show floating
12. Show tissue
13. Say lamp
14. Show bandage
15. Say painting
16. Say milk
17. Show arguing
18. Show vehicle
19. Say star
20. Show discussing
21. Show experimenting
22. Say escalator
23. Say globe
24. Say compass
25. Say hinge

Basic Skills Subtest

1. Point to the biggest stick
2. Point to the smallest stick
3. Point to the biggest ball
4. Point to the smallest ball
5. Count 3 bears
6. Show 1
7. Count 5 squirrels
8. Say 5
9. Say A
10. Show k
11. Say 7
12. Show 22
13. Point to me
14. Point to to
15. Say off
16. Point to cold
17. Add 3 + 4 = 7
18. Point to cereal
19. Say school
20. Show couple
21. Show ½ of 6 = 3
22. Say phone
23. Say knife
24. Show honest
25. Say 1120

Appendix C
Dollaghan and Campbell's (1998) Nonwords

1. /naɪb/
2. /voʊp/
3. /taʊʤ/
4. /dɔɪf/
5. /teɪvak/
6. /ʧoʊvæg/
7. /væʧaɪp/
8. /nɔɪtaʊf/
9. /ʧinɔɪtaʊb/
10. /naɪʧoʊveɪb/
11. /dɔɪtaʊvæb/
12. /teɪvɔɪʧaɪg/
13. /veɪtaʧaɪdɔɪp/
14. /dævoʊnɔɪʧig/
15. /naɪʧɔɪtaʊvub/
16. /tævaʧinaɪg/

Appendix D
CELF-P2 Sentences

1. He is nice.
2. They play with blocks.
3. The boy fell and hurt himself.
4. Didn't the boys eat the apples?
5. The rabbit was not put in the cage by the girl.
6. Was the teacher followed by the children?
7. The big, brown dog ate all of the cat's food.
8. The play castle was built by the girls and boys.
9. The kindergartner cannot cross the street by himself.
10. Because tomorrow is Saturday, we can stay up late tonight.
11. The toy was not returned to the shelf by the girl.
12. The dad bought a book for his son who likes funny stories.
13. The girl who won the prize at the school party was very excited.

REFERENCES

ASHA. (2008). *Incidence and Prevalence of Communication Disorders and Hearing Loss in Children – 2008 Edition.* Rockville, MD: Author.

ASHA. (2008). *Roles and Responsibilities of Speech-Language Pathologists in Early Intervention: Guidelines.* Rockville, MD: Author.

American Guidance Service. (1990). *AGS Early Screening Profiles.* San Antonio, TX: Pearson.

American National Standards Institute. (1991). *ANSI.* Washington, DC.

Archibald, L. & Gathercole, S. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, *49(5)*, 970-983.

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

Bernthal, J. E., Bankson, N. W., & Flipsen, P. (2012). *Articulation and phonological disorders.* Boston, Mass.; London: Allyn & Bacon.

Bishop, D., Adams, C., & Norbury, C. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain, and Behavior, 5,* 158-169.

Chen, Y. (2011). Identification of articulation error patterns using a novel dependence network. *IEEE Transactions on Biomedical Engineering , 58,* 3061-3068.

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

Cronan, S., Blanchet, P., Klonsky, B., & Piazza, N. (2014). The effect of modeled /r/ articulatory disorder on listener perceptions of speech skills and personality traits. *Contemporary Issues in Communication Sciences and Disorders, 41,* 169-178.

Deevy, P., Weil Wisman, L., Leonard, L., & Goffman, L. (2010). Extending the use of the NRT to preschool-age children with and without specific language impairment. *Language, Speech, and Hearing Services in Schools, 41,* 277-288.

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

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent syllable effects in a nonsense-word repetition task. *Journal of Speech and Hearing Research, 36,* 1051-1054.

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research: JSLHR, 43(4)*, 865–878.

Eisenberg, S. & Guo, L. (2013). Differentiating children with and without language impairment based on grammaticality. *Language, Speech, and Hearing Services in School, 44,* 20-31.

Flipsen, P. (2006). Measuring the intelligibility of conversational speech in children. *Clinical Linguistics & Phonetics, 20,* 303-312.

Francois, J., Coufal, K., & Subramanian, A. (2015). Student preparation for professional practice in early intervention. *Communication Disorders Quarterly, 36,* 177-186.

Gathercole, S. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(04), 513–543.

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

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

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation, 2nd ed.* Minneapolis, MN: Pearson.

Graf Estes, K., 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: JSLHR, 50(1),* 177–195.

Gravetter, F. & Wallnau, L. (2013). *Statistics for the behavioral sciences.* Belmont, CA: Cengage Learning.

Grunwell, P. (1987). *Clinical phonology* (2nd ed.). London: Croom.

Hodson, B. (2004). *Hodson Assessment of Phonological Patterns, 3rd ed.* Austin, TX: Pro-Ed.

Johnson, C. A., Weston, A. D., & Bain, B. A. (2004). An objective and time-efficient method for determining severity of childhood speech delay. *American Journal of Speech-Language Pathology / American Speech-Language-Hearing Association*, *13*(1), 55–65.

McLeod, S. & Harrison, L., (2009). Epidemiology of speech and language impairment in a nationally representative sample of 4- 5-year-old children. *Journal of Speech, Language and Hearing Research: JSLHR, 52(5),* 1213-1229.

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research: JSLHR*, *46*(5), 1138–1153.

Montgomery, J., Evans, J., & Gillam, R. (2009). Relation of auditory attention and complex sentence comprehension in children with specific language impairment: A preliminary study. *Applied Psycholinguistics, 30(1),* 123-151.

Nash, H., Leavett, R., & Childs, H. (2011). Evaluating the GAPS test as a screener for language impairment in young children. *International Journal of Language and Communication Disorders*,*46(6)*, 675-685.

Newton, C. (2012). Between-word processes in children with speech difficulties: insights from a usage-based approach to phonology. *Clinical Linguistics & Phonetics, 26(8),* 712-727.

Paul, R., & Norbury, C. (2012). *Language disorders from infancy through adolescence: listening, speaking, reading, writing, and communicating*. St. Louis, Mo.: Elsevier.

Sahlen, B., Reuterskiold-Wagner, C., Nettelbladt, U., & Radeborg, K. (1999). Non-word repetition in children with language impairment-pitfalls and possibilities. *International Journal of Language & Communication Disorders, 34(3)*, 337–352.

Semel, E., Wiig, E., & Secord, W. (2004). *Clinical Evaluation of Language Fundamentals – Preschool, 2nd ed.* San Antonio, TX: Pearson.

Shriberg, L. & Kwiatkowski, J. (1982b). Phonological disorders III: a procedure for assessing severity of involvement. *Journal of Speech and Hearing Disorders, 47,* 256-270.

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research*, *36*(1), 105–140.

Shriberg, L., Austin, D., Lewis, B., McSweeney, J., & Wilson, D. (1997a). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40,* 708-722.

Shriberg, L., Lohmeier, H., Campbell, T., Dolloghan, C., Green, J., & Moore, C. (2009). A nonword repetition task for speakers with misarticulations: The syllable repetition task (SRT). *Journal of Speech, Language, and Hearing Research, 52,* 1189-1212.

Tomblin, J., Records, N., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245-1260.

Tyler, A. A., Miccio, A. W., Hoffman, P. R., Hodson, B., & Bleile, K. (2002). Five views of the elephant: Perspectives on the assessment of articulation and phonology in preschoolers. *American Journal of Speech-Language Pathology*, *11*, 213–214.

Vance, M., Stackhouse, J., & Wells, B. (2005). Speech-production skills in children aged 3-7 years. *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *40*(1), 29–48.

Williams, L., McLeod, S., & McCauley, R. (2010). *Intervention for Speech Sound Disorders in Children.* Baltimore, MD: Paul H. Brookes Publishing Co.