USING NARROW AND GLOBAL MEASURES TO IDENTIFY CHILDREN WITH LI IN A POPULATION SAMPLE

A Master's Thesis

Presented to

The Faculty of the Department

of Communication Sciences and Disorders

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

By

Tiana M. Cowan

May 2017

USING NARROW AND GLOBAL MEASURES TO IDENTIFY CHILDREN WITH LI IN A POPULATION SAMPLE

Tiana Cowan, B.S.

APPROVED:

Anny Castilla-Earls, Ph.D., CCC-SLP
University of Houston
Committee Chair

Ferenc Bunta, Ph.D., CCC-SLP
University of Houston
Committee Member

Michelle Ivey, Ph.D., CCC-SLP
University of Houston
Committee Member

Kristi Santi, Ph.D.
University of Houston
Committee Member

Antonio D. Tillis, Ph.D.
Dean, College of Liberal Arts and Social Sciences
Department of Hispanic Studies

USING NARROW AND GLOBAL MEASURES TO IDENTIFY CHILDREN WITH LI IN A POPULATION SAMPLE

An Abstract of a Master's Thesis

Presented to

The Faculty of the Department

of Communication Sciences and Disorders

University of Houston

In Partial Fulfillment

Of the Requirements for the Degree of

Master of Arts

By

Tiana M. Cowan

May 2017

Using Narrow and Global Measures to Identify Children with LI in a Population Sample

Abstract

Appropriate identification of children with language impairment (LI) is a persistent clinical issue in the field of communication disorders. Supplemental measures derived from the clinical markers of LI have been evaluated for diagnostic accuracy and clinical utility to improve diagnostic practices. Three of such measures are the percent of grammatical communication units (PGCU), the finite verb morphology composite (FVMC) and errors per communication unit (ECU). Although there is promising initial data to support the use of these supplemental measures, they have only been evaluated by two-gate studies to date, which creates spectrum bias. Spectrum bias is the methodological flaw with the greatest potential to overestimate the diagnostic accuracy of a measure (Dollaghan & Horner, 2011; Lijmer et al., 1999; Rutjes, et al., 2005). Therefore, this one-gate study aims to evaluate the diagnostic accuracy and clinical utility of the supplemental measures in a participant pool that reflects a clinical population. Participants included 141 monolingual children ages four through six recruited from a public school in upstate New York. Findings indicate that PGCU, FVMC, and ECU do not have acceptable diagnostic accuracy to identify LI in children ages four through six. Findings from this study may contribute to recommendations for identification of LI.

Acknowledgments

I would like to thank Dr. Anny Castilla-Earls for her tireless support and guidance throughout this process. Learning how to think and write like a scientist is a life-long process that I have been lucky enough to embark on with the mentorship of Dr. Castilla. Your patience, and kindness helped me be kinder to myself throughout the creation of this Master's thesis. Thank you for your hard work and mentorship.

Table of Contents

List of Tables

Introduction

Appropriate identification of children with language impairment (LI) is a persistent clinical issue in the field of communication disorders (Guo & Schneider, 2016; Rice & Wexler, 1996). LI is characterized by language deficits in the absence of other developmental disabilities. Deficits may be in the areas of syntax, morphology, semantics, phonology, and/or pragmatics. Although the profile of language deficits may vary from individual to individual, delayed acquisition of finiteness markers (e.g., verb tense markers and agreement) is the hallmark clinical marker of LI (Rice & Wexler, 1996). Therefore, children with LI produce utterances that are less grammatical than those produced by their age-matched peers (Leonard, 2014; Moore, 2001). These language deficits begin in early childhood and continue into adolescence and adulthood (Betz, Poll, & Miller, 2010).

Although the clinical markers of LI have been well documented, identifying children with LI continues to prove challenging in the field of communication sciences and disorders. The common practice of using standardized testing for identification of LI is problematic since many of the standardized measures available either do not report or lack adequate diagnostic accuracy (Plante & Vance, 1994; Spaulding, Plante, & Farinella, 2006). According to Spaulding, Plante and Farinella (2006) identification accuracy (e.g., how accurate a measure to identify the disorder is?) is measured by sensitivity and specificity data. Sensitivity measures diagnostic accuracy by calculating the percentage of children with language impairment that are correctly identified as impaired. Specificity measures diagnostic accuracy by calculating the percentage of typically developing children correctly identified as typical (Plante, Spaulding, & Farinella, 2006). Acceptable sensitivity and specificity percentages range from 80-89% correctly identified children. Good sensitivity and specificity percentages include $\geq$90% correctly identified children.

Plante and Vance first advocated the use of sensitivity and specificity measurements to ensure adequate diagnostic accuracy in 1994. Despite this recommendation, Spaulding et al (2006) found that only nine out of 43 standardized measures used for the assessment of children reported sensitivity and specificity information. Of the nine assessments, only five reported acceptable (i.e., 80-89% accuracy), or good (i.e., ≥90% accuracy) diagnostic accuracy: the Clinicial Evaluation of Language Fundamentals-Fourth Edition (Semel, Wiig & Secord, 2003), Preschool Language Scales-Fourth Edition (Zimmerman, Steiner, & Pond, 2002), Test of Early Grammatical Impairment (Rice & Wexler, 2001), Test of Language Competence-Expanded Edition (Wiig & Secord, 1989), and Test of Narrative Language (Gillam & Pearson, 2004; Spaulding, et al., 2006).

Due to the persistent clinical issues in identifying children with LI via the use of standardized assessments, studies have investigated supplemental grammatical measures to increase diagnostic accuracy. These measures are based on the clinical markers of LI, and in some cases have demonstrated better diagnostic accuracy than currently used standardized measures (Aram, Morris & Hall, 1993; Ebert & Scott, 2014; Eisenberg & Guo, 2013). Both narrow and global measures of grammaticality have been evaluated for identification potential. Narrow measures analyze specific errors made within an utterance. For example, analyzing errors in verb tense agreement, or pronoun usage would be two examples of narrow measures (Souto et al., 2014). Global measures examine the overall grammaticality of a child's language production across utterances and do not focus on the specific error patterns. Any combination of errors in either morphology, syntax, or semantics may be analyzed on the overall, or global, impact the deficits have on grammaticality (Colozzo et al., 2011; Fey, et al., 2004). Due to the potential for supplemental language measures to have better diagnostic accuracy than the current standardized language assessments, it is important to further investigate supplemental measures for the identification of children with LI (Dunn, Flax, Sliwinski, & Aram, 1996; Eisenberg & Guo, 2013).

Identifying Children with LI Using Global Measures

Global measures of grammar evaluate overall, or global, grammaticality across utterances. One global measure that has been evaluated by studies includes the percent of grammatical communication units (C-Units; PGCU). A C-Unit is an independent clause, plus all of its dependent clauses (Loban, 1976). The PGCU calculates a percentage of grammaticality by identifying C-Units with grammatical errors, as compared to the C-Units produced correctly. The PGCU is based on the Developmental Sentence Scoring (DSS) sentence point measure (Lee, 1974; Souto, Leonard, & Deevy, 2014). The sentence point in the DSS analysis evaluates the grammaticality of 50 sentences in language samples based on the correct use of grammatical morphemes, sentence structure, and semantics. All grammatical sentences earn a score of one. All ungrammatical sentences earn a score of zero (Lee, 1974). The PGCU procedures were derived from the DSS analysis (Eisenberg & Guo, 2016; Souto et al., 2014)

Previous studies have demonstrated that PGCU has acceptable diagnostic accuracy for preschool age children, as well as school age children (Eisenberg & Guo, 2013; Eisenberg & Schneider, 2016; Souto, et al., 2014). Eisenberg and Guo (2013) compared the diagnostic accuracy of three global measures of grammar for identifying LI in three year olds: PGCU (referred to in the study as PGU), percentage sentence point (PSP), and percentage verb tense agreement (PVT). Of the three measures of grammaticality, PGCU had the greatest diagnostic accuracy. A cutoff score of 58% PGCU showed good sensitivity (100%) and acceptable specificity (88%) (Eisenberg & Guo, 2013). Similarly, Souto et. al. (2014) evaluated the diagnostic accuracy of PGCU in differentiating four and five year olds with and without LI. They found that PGCU had good diagnostic accuracy (e.g., 100% sensitivity, 100% specificity), although cutoff scores were not reported, which limited the clinical applicability of the study. (Eisenberg & Schneider, 2016; Souto, et al., 2014). Although there is promising initial evidence to support the diagnostic use of PGCU, the measure's diagnostic accuracy has yet to be examined for ages four, and five with reported cutoff scores.

Identifying Children with LI Using Narrow Measures

One significant clinical marker of LI throughout childhood and adolescence is morphosyntactic difficulty, including deficits in finite verb morphology (Colozzo et al., 2011; Fey, et al., 2004; Poll, Betz, & Miller, 2010). Preschool age children with LI produce fewer verb tenses and use tense agreement morphemes more inconsistently than age-matched TD peers (Cleave & Rice, 1997; Conti-Ramsden, Botting, & Faragher, 2001). Therefore, studies have examined the diagnostic accuracy of tense measures of finite verb morphology. One composite measure developed to differentiate LI from TD children based on productivity of tense agreement morphemes is the Finite Verb Morphology Composite (FVMC). The FVMC computes the percentage of correct use of third person singular -s, past tense -ed, and copula and auxiliary be in spontaneous production contexts (Bedore & Leonard, 1998). Studies have demonstrated acceptable diagnostic accuracy of the FVMC for the identification of LI in preschool children (Bedore & Leonard, 1998; Guo & Eisenberg, 2013; Guo & Eisenberg, 2014). For example, Bedore and Leonard (1998) investigated the differentiating power of the FVMC. Participants included 19 children with specific language impairment (SLI), and 19 typically developing peers, ages three years eight months to five years seventh months. The FVMC was derived from a spontaneous play-based language sample. Bedore and Leonard (1998) found acceptable diagnostic accuracy for the identification of SLI in the preschool age children. However, when combining the FVMC with the additional measure of mean length utterance (MLU) the diagnostic accuracy improved. When applied to school age children, the FVMC had inconsistent diagnostic power (Moyle, Karasinski, Weismer, & Gorman, 2011; Souto, et al., 2014). Souto et. al. (2014) found that the FVMC had good diagnostic accuracy for differentiating four and five year olds with and without LI and Gladfelter and Leonard (2013) found good diagnostic accuracy for young school age children, ages five to five years six months. Gladfelter and Leonard (2013) reported good accuracy when applying a cutoff FVMC score of 85%. Moyle et al. (2011) found that the FVMC was not a clinically viable tool since it under-identified 50% of children (mean age of seven years nine months) in their study. Moyle believed that this decrease in diagnostic power was due to the mastery of tense agreement morphemes by older children with LI (Moyle, et al., 2011). However, the measure appears to be diagnostically valuable for children up to age six (Eisenberg & Guo, 2013; Eisenberg & Schneider, 2016; Souto, et al., 2014).

Although the FVMC has demonstrated acceptable diagnostic accuracy for the identification of LI in children, Eisenberg and Guo (2016) explored the diagnostic value of a new narrow measure, errors per communication unit (ECU). ECU identifies the total number of grammatical errors (e.g., related to nouns or verbs) across communication units (C-Units) (Eisenberg & Schneider, 2016; Scott & Windsor, 2000). However, some studies have included a larger variety of utterances that do not fit this strict definition of a C-Unit (e.g., utterances omitting auxiliary be verbs) (Guo & Schneider, 2016). Once all errors are tallied, the total is divided by the number of C-Units. Both ECU and PGCU were employed by Guo and Schneider (2016) in a recent study. The participants included 61 six year olds (50 TD, and 11 with LI), and

67 eight year olds (50 TD, and 17 with LI). Both measures were derived from a narrative generation task. ECU demonstrated acceptable diagnostic accuracy up until age six, while PGCU demonstrated acceptable diagnostic accuracy until age eight (Eisenberg & Schneider, 2016). The promising initial data about the diagnostic accuracy of ECU warrants further examination by future studies, including larger age groups, and different language elicitation methods to evaluate the measure's clinical diagnostic potential use.

Threats to Diagnostic Accuracy

Although there is some promising initial data for the differentiating ability of narrow and global measures, there is a need to examine the diagnostic accuracy with a stricter study design than has previously been completed. The two primary study designs that examine diagnostic accuracy are one-gate and two-gate (Dollaghan & Horner, 2011). One-gate designs endeavor to create a sample that represents the general population (e.g., a population sample). Therefore, participants are recruited without needing to disclose the presence or absence of disease/disorder. Conversely, a two-gate study recruits participants based on the presence or absence of disease/disorder in order to create case and control groups. Consequently, two-gate studies create a spectrum bias wherein the case-controlled groups represent extremes in terms of language ability (e.g., severely language impaired, or above average language ability). Spectrum bias is the methodological flaw with the greatest potential to overestimate the diagnostic accuracy of a measure (Dollaghan & Horner, 2011; Lijmer et al., 1999; Rutjes, et al., 2005). To date, the supplemental measures' diagnostic accuracy for the identification of LI have only been examined by two-gate studies. Therefore, the measures have yet to be tested in a population sample. Since a population sample is one that is truly representative of the general population, the majority of participants should have average language abilities. In a two-gate study, there tends to be a dearth of participants with average language ability. This is important because it implies that the reported diagnostic accuracy for the supplemental language measures may be overestimated for clinical application (Dollaghan & Horner, 2011).

Although supplemental measures have only been evaluated by two gate studies, Guo and Schneider (2016) designed a study that endeavored to account for spectrum bias. The investigators created subgroups within the TD, and LI participants by asking the teachers from a public school to refer children to the study that were high achieving, average achievement, and low achieving. Therefore, theoretically, ensuring a sample that represented the distribution of language abilities found within the public schools. The study evaluated the diagnostic accuracy of the FVMC, PGCU (referred to as PGU in the study), and ECU for six and eight year olds. All measures demonstrated good diagnostic accuracy at age six. At age eight only PGCU demonstrated good sensitivity and specificity (Guo & Schneider, 2016). The present one gate study will evaluate the supplemental measures' diagnostic utility for four, five and six year olds in a clinical population.

Purpose of This Study

The purpose of the present study is to evaluate the diagnostic accuracy of supplemental measures derived from a story retelling to identify children with LI in a population sample. The current evidence suggests that narrow and global measures of grammaticality may identify children with LI with acceptable sensitivity and specificity (Conti-Ramsden et al., 2001; Eisenberg & Guo, 2013; Eisenberg & Schneider, 2016; Souto, et al., 2014). However, the diagnostic accuracy is unknown when differentiating TD from LI children in a population sample. Therefore, the diagnostic accuracy of these measure must be evaluated in one-gate design studies. As aforementioned, two-gate study designs can create a spectrum bias, since the participant groupings tend to include children that are either severely language impaired, or above average in terms of language ability (Dollaghan, 2007). Therefore, the measures may appear more sensitive or specific at identifying language impairment than might be true in an actual clinical setting (e.g., a setting with children with average language abilities, or a mild language impairment) (Lijmer et al., 1999; Rutjes, et al., 2005). This study aims to answer the question: do narrow and global morphosyntactic measures derived from a story-retell task show adequate diagnostic accuracy to identify LI in a population sample?

Methods

Participants

The present study involves a secondary data analysis. All data were collected for the study “Diagnostic Accuracy of the CELF-P2 and PLS-5” (Castilla-Earls, under review). Participants were recruited from one school district in upstate New York. All children within the age range of four years to six years eleven months within the school district were invited to participate in the study. The goal was to create a sample that was representative of the language abilities of the entire school district's population by inviting all children from all classrooms. One hundred and seventy children were recruited to participate in the original study. Prior to the evaluation, all participants passed a cognitive screening, and a hearing screening. Each child scored ≥80 on the Non-Verbal Scale of the Kaufman Brief Intelligence Test 2 (K-BIT 2; Kaufman & Kaufman, 2004). Hearing was screened by otoacoustic emission (OAE) at 1,000 to 4,000 Hz. If a participant did not pass the OAE screening the first time, a second one would be administered to verify the negative results. If the child did not pass two OAE screenings, then a pure tone screening was administered to determine hearing status. Exclusionary criteria included not passing a hearing screening (n=5), not passing a cognitive screening (n=7), more than 20%

exposure to a second language at home (n=1), turning seven during the study (n=5), not providing consent (n=5), withdrawn by parent (n=1), changing schools during the evaluation period (n=1), diagnosis of autism spectrum disorder (n=1), and testing could not be scheduled within a three week period (n=3). The remaining participant pool (n=141) included 36 four year olds (Mean age= 54 months; SD= 3 months; range= 48 to 59 months), 69 five year olds (Mean age= 5;8; SD= 3 months; range= 60 to 71 months) and 36 six year olds (Mean age = 77 months; SD= 4 months; range= 72 to 83 months). The sample was predominantly comprised of white (94%), non-Hispanic (90%) children. All children's parents completed a survey and indicated languages spoken in the home. All children invited to participate in the final study were monolingual English speaking children, introduced to other languages less than 20% of the time at home. Across age groups, males comprised 45% of the sample, and females comprised 55%. The parents reported the highest level of education achieved by the mother to be middle school (~1%), high school (15%), some college (19%), Associate's' degree (13%), Bachelor's degree (26%), and Graduate degree (27%).

Reference Measure

The Structured Photographic Expressive Language Test, 3rd edition (SPELT-3; Dawson, Stout, & Eyer, 2003) was administered as a part of the original study, and the standard scores were utilized as the reference measure in the present study. The SPELT is comprised by a booklet of photographs and a set of questions designed to elicit specific morphological and syntactic structures. The administration time is approximately 15 minutes. This measure is referred to as the gold standard for identification of language impairment in children, and has been verified in independent research studies (Greenslade, Plante, & Vance, 2009; Plante & Vance, 1995). Following the recommendation of Reilly et al. (2014) a cutoff score of 82, which is 1.25 SDs below the mean, was applied to identify children with LI in the present study. When differentiating TD and LI groups using a standard score (SS) of 82, 13 children were identified as having LI, and 131 were identified as TD in this group of children.

Procedures

All data were collected in one session. The testers administered the cognitive screening, hearing screening, the SPELT, and collected a language sample. The language sample was collected using a narrative retell task. The narrative retell task utilized Frog Goes to Dinner, a wordless picture book by Mercer Mayer (1974),. The story is about a boy going to dinner with his family and a frog. First, the examiner told the story of the child going to dinner and the frog causing a problem in the restaurant. Then, the child was directed to retell the same story using only the pictures as a guide. The children were prompted as necessary to continue to retell the story and elicit adequate language (e.g., “What is he doing here?”). To avoid subject bias, testers, transcribers, and coders were blind to the language status of the child.

Measures

Errors/C-Unit

Eisenberg and Guo (2013) and Guo and Schneider’s (2016) coding scheme was applied in this study. C-Units that omitted subjects, and auxiliary be verbs were included in the Eisenberg and Guo (2013) found that the grammaticality analyses had more differentiating power for three year olds when these C-Units (e.g., without subjects) were analyzed. Nonclausal utterances that express complete thoughts were also included. The most frequent context for a nonclausal utterance to be produced was in response to a question. For example, if asked, “What are they doing in this picture?” The child could appropriately answer, “eating and drinking.” This nonclausal utterance would be analyzed for grammaticality, and errors.

Errors were coded according to the following criteria:

(a) Tense errors were defined as any omission or incorrect use of 3SG –s, regular past tense –ed, copula and auxiliary be, auxiliary do, irregular past tense, and irregular third person verb forms. Any inappropriate production of an infinitive verb was counted as a tense error; however, only in the case that the sentence clearly obligated a tense marker.

(b) Pronoun errors were defined as: substitution errors for subject pronouns, object pronouns, reflexive pronouns, possessive pronouns, and possessive determiners, and omissions or incorrect uses of relative pronouns. Reference errors were included in this category.

(c) Grammatical morpheme errors were operationally defined as omission or incorrect uses of grammatical morphemes other than those included in the aforementioned categories of tense marking, and pronouns. Examples of such morpheme errors include: plural -s, prepositions, and present and past participles.

(d) Argument structure errors were operationally defined as omissions of objects in any context surrounding verbs (e.g., They go to.).

(e) Other errors were operationally defined as any other syntactic error that did not fit the any of the aforementioned prescribed categories (e.g. omission of possessives).

All codes for errors are displayed in Table 1. After the errors were coded, the total number of errors was calculated. The number of Errors Per C-Unit was calculated by tallying the total errors/the total utterances produced to find the average of errors per utterance (Guo & Eisenberg 2013; Gladfelter & Leonard 2013).

PGCU

PGCU is a global measure that evaluates overall grammaticality of utterances produced. It calculates a percentage of C-Units that were produced grammatically correctly. Utterances missing a copula be verb (e.g., She sad) and non-imperative utterances without a subject (e.g., Want a cookie) were included in the analysis. Utterances were marked ungrammatical for any of the aforementioned error codes listed under Errors Per C-Unit, including semantic errors. All grammatical codes listed in Table 1.

FVMC

The finite verb morphology composite (FVMC) computes the percentage of correct use of third person singular -s, past tense -ed, and copula and auxiliary be. It calculates the percentage of correct use in obligatory contexts, as defined by the percentage of times the grammatical morphemes are used in the utterances that need the markers in order to be considered correct. Errors will only be marked when the obligatory context is clear (e.g., a clear statement of time, or a preceding verb tense). Analyzed C-Units must have a subject and a verb, unless the omitted verb is an auxiliary be verb. This is so because a subject is necessary to create an obligatory context for tense agreement (Eisenburg & Guo, 2013; Eisenburg & Schneider, 2016). Any irregular past tense verbs, and auxiliary do verbs were excluded from analysis, in accordance with Bedore and Leonard (1998). All FVMC codes displayed in Table 1.

MLU

MLU was calculated by segmenting the language transcripts into C-Units then analyzing the transcriptions by the Systematic Analysis of Language Transcripts (SALT). To determine MLU, SALT calculated the average number of morphemes each C-Unit contained. Frequently, measures of grammaticality and finite morphology depend upon production errors in morphemes, as compared to omission errors. If a child does not produce enough language to analyze for errors, then it could be a contributing factor to under identification. Therefore, it is important to examine MLU as a dependent variable (Bedore & Leonard, 1998).

Procedures

The narrative samples were transcribed and segmented in C-Units by trained research assistants (SALT; Miller & Chapman, 2000). The primary author of this graduate thesis coded all errors using the coding system described before. Due to the demographics of the participants of this study, all errors were coded according to the rules of Standard American English (SAE). SAE is the primary variety of English typically used by middle and upper socioeconomic status Caucasian families, and is taught in American classrooms (Beyer & Kam, 2012). If a child produced a word that was thought to be a dialectal variation, then the SPELT III manual, and Dr. Castilla-Earls were consulted to determine if the variation produced was an error or a dialectal difference. After coding and analyzing the transcripts in SALT, all error codes were exported to an Excel database. Basic equations were programmed in excel to calculate the percentage of correct FVMC verb tense productions, the percentage of grammatical C-Unit productions, and the total number of Errors Per C-Unit.

Reliability

Inter-rater reliability for transcription and coding of the language samples was calculated for 20% of the samples. The primary author of this thesis trained a graduate assistant from the State University of New York at Fredonia on all error codes (displayed in Table 1) to complete reliability coding in SALT. The primary author reviewed all transcripts coded for reliability line by line to verify agreement. When there was a discrepancy in coding, Castilla-Earls was consulted to assist in the reconciliation of the disagreement. All supplemental measures had over 90% inter-rater reliability for coding. The percentage of inter-rater agreement for grammaticality coding was 97%, for Errors Per C-Unit was 92%, and for FVMC errors was 95%.

Analyses

The purpose of this study was to examine the diagnostic accuracy of narrow and global measures derived from a story-retell task to identify LI in a population sample. To this end, two different sets of analysis were conducted. First, analyses of group differences on FVMC, PGCU, and ECU between age groups and LI and TD children were conducted. Since there are not specified cutoff scores to apply for FVMC, PGCU and ECU for narrative retell language samples collected from four to six year olds, cutoff scores were calculated from the group variance to estimate which children showed low performance. In addition, group differences between TD and LI children were calculated to examine if the FVMC, PGCU and ECU can detect differences in performance between the groups. Second, diagnostic accuracy analyses were conducted with various cutoff scores to examine the sensitivity and specificity of the FVMC, PGCU and ECU in a population sample. Cutoff scores for the supplemental measures were tested at 1, 1.25, and 2 SD from the mean of the age group.

To determine the diagnostic accuracy of the supplemental measures sensitivity, specificity, and likelihood ratios were calculated for each measure (Dollagan 2007). Acceptable sensitivity considered to be between 80 and 89% identification of true disorder. Good sensitivity is considered to be ≥90% identification of true disorder (Plante & Vance, 1994). Likelihood ratios (LR) were computed from the levels of sensitivity and specificity. A positive LR looks at ratios of true LI to false LI. Negative LR calculates ratio of false TD to true TD. According to Dollaghan (2007) acceptable LR + range from 5.00 and 9.99, and good LR + are ≥ 10.00. Acceptable LR - range from .11 and .20, and good is ≤.10. Specificity, sensitivity and likelihood ratios were calculated by the Evidenced-Based Centre in Toronto (http://ktclearinghouse.ca /cebm/practise/ca/calculators). All information is reported and analyzed below.

Results

A significant omnibus F-test for a one-way ANOVA analyzed the differences in age group performance for the FVMC, PGCU, ECU, total number of utterances, obligatory contexts for the FVMC tenses, and the MLU in morphemes, displayed in Table 3. There were no statistically significant differences in performance between age groups for the FVMC, F(2,138) = 1.55, p= .217, PGCU, F(2,138) = 1.59, p= .208, or EGU, F=(2,138)= .184, p=.160. However, there were statistically significant differences between age groups for MLU in morphemes, F(2,138) = 13.49, p = .000, and obligatory contexts, F=(2,138) = 9.93, p = .000. Post hoc analyses, provided in Table 4, indicated that five year old children had higher MLU in morphemes than four year olds, (mean=7.64, 6.45; SD=1.55, 1.82; p =.002) and higher number of FVMC obligatory contexts, (mean = 19.22, 24.52; SD = 7.99, 9.33; p=.012). No statistically significant differences were found between five and six year olds.

A F-test analyzed differences in performance between LI and TD groups on FVMC, PGCU, ECU, MLU, and obligatory contexts, results presented in Table 5. There were statistically significant differences in performance between LI and TD children on FVMC, ECU, PGCU, MLU, and obligatory contexts. Across all supplemental measures the TD group performed statistically significantly better than the LI group. On the FVMC TD participants performed statistically significantly better than LI participants, F=(126,13)=80.85, p=.000. The Additionally, TD participants scored an average of 22.69 percentage points better than LI participants. Interestingly, the SD of performance for LI children (SD=5.98) was approximately five times greater than the SD for TD children (SD=28.72), indicating a much wider variety of scores achieved by LI participants as compared to TD. On PGCU TD participants scored statistically significantly better than LI participants, F=(126,13)=27.69, p=.000. In addition, TD participants scored an average of 24.13 percentage points higher than LI participants. Again, the LI participants had a much greater SD of performance (SD=23.69) than TD children (10.74). On ECU TD participants produced statistically significantly less errors than LI participants, F=(126, 13)=23.67, p=.000. In fact, TD participants produced an average of .314 less errors per utterance than LI participants. Once more, the LI group had a larger SD (SD=.31) than the TD group (SD=.17). A t-test was conducted to compare MLU in morphemes and obligatory contexts between LI and TD groups. Results indicated that the TD group (mean=7.641, SD =1.734) produced significantly longer utterances than the LI group (mean = 5.872, SD = 1.896); t (137)= 3.471, p=.000. Additionally, the TD group (mean=24.920, SD= 9.302) produced significantly more verb tenses than the LI group (mean=17.000, SD= 8,889); t (137)=2.934, p=.000.

Sensitivity and Specificity

Cutoff scores for FVMC, PGCU, and ECU were calculated 1, 1.25 and 2 SD from the mean to estimate the diagnostic ability to identify which children showed low performance. Table 6 presents the cutoff scores and associated indices of diagnostic accuracy for FVMC. Table 7 presents the cutoff scores and associated indices of diagnostic accuracy for PGCU. Table 8 presents the cutoff scores and associated indices of diagnostic accuracy for ECU. ECU, FVMC and PGCU, utilizing cutoff scores derived from 1 SD from the mean demonstrated unacceptable sensitivity, between 53.8% and 76.9%. The FVMC demonstrated the poorest sensitivity, identifying 53.8% (n=7) of the children with LI. ECU was the most sensitive of the measures, and approximated acceptable diagnostic accuracy. ECU demonstrated 76.9% sensitivity, and 96.1% specificity for differentiating children with LI from TD between the age of four and six when utilizing cutoff scores that were 1 SD from the mean. The LR + was 20.15, which means those with LI are 20.15 times more likely to obtain a fail score than a child with TD. The LR- was .24, which demonstrates that children with LI were .24 times as likely to obtain a pass score when compared to TD peers. All three supplemental measures showed good specificity for the accurate identification of TD status (over 90%).

Utilizing cutoff scores that were 1.25 SD from the mean decreased sensitivity for FVMC, PGCU, and ECU to approximately 50% sensitivity, meaning that accurate identification of LI status was approximately at chance. FVMC sensitivity decreased to 46.15% (n=6), meaning that it under identified over 50% of LI participants utilizing cutoff scores 1.25 SD from the mean. All three measures demonstrated good specificity for the accurate identification of TD status (over 90%).

Although the SPELT III manual recommends identifying LI 1 SD below the mean, and Reilly (2014) recommends identifying LI as 1.25 SD below the mean, applying these cutoff scores yielded unacceptable diagnostic accuracy for the FVMC, PGCU, or ECU. Therefore, other possibilities were explored in terms of both reference measure and supplemental measure cutoff scores to investigate the impact on diagnostic accuracy. Souto et al.'s (2014) inclusionary criteria for participants with LI included a score in the first percentile rank on the SPELT II, the most current version of the evaluation at the time of data collection. Therefore, this present study applied similar diagnostic criteria, and identified participants with LI as obtaining scores in the first percentile. Utilizing the more restrictive diagnostic criteria to identify LI, only three children were diagnosed as LI in this study. Next, cutoff scores 2 SD from the mean on the supplemental measures were applied to evaluate the diagnostic accuracy. By restricting the diagnostic criteria, and applying cutoff scores on the FVMC, PGCU, and ECU that were 2 SD below the mean, sensitivity and specificity dramatically improved for all supplemental measures, results displayed in Tables 6 through 8. PGCU and ECU demonstrated good diagnostic sensitivity at 100% by correctly identifying all three children as LI. Specificity remained good (99%; 97) as well. The

FVMC continued to demonstrate unacceptable sensitivity (67%) and good specificity (97%).

Discussion

The present graduate thesis evaluated the diagnostic accuracy of the FVMC, ECU, and PGCU for differentiating young school aged children with and without LI between the ages four and six. The results indicated that FVMC, ECU, and PGCU have low levels of sensitivity for identifying with LI when following the reference measure manual's recommendation of diagnosing LI 1 SD below the mean, and by following the recommendation of Reilly (2014) and LI as 1.25 SD below the mean. When diagnostic criteria for identifying LI on the reference measure was more restrictive, in alignment with Souto et al. (2014), the supplemental measures showed improved diagnostic accuracy using cutoff scores 2 SD from the mean.

In general, the TD participants performed better across all measures than the LI group, displayed in Table 5. The mean FVMC score from TD participants was 23% higher than the LI participants' mean score (96.09; 73.39%), indicating a pattern that those with LI do struggle with finiteness marking when compared to their peers. However, the SD for performance was 28.72 from the LI group, indicating a very wide variation in performance. This trend was observed in the other supplemental measures as well. For example, the TD group's mean score on the PGCU was 24% higher than the LI group's mean score. However, the LI group's SD was much wider than the TD group's (23.69; 10.74). In ECU the TD group made an average of .17 Errors Per C-Unit, and the LI group made .48. However, the LI group's SD was .31, as compared to the TD group's SD of .14. Therefore, although there were significant group differences in performance, the variation in performance within the LI group decreased the diagnostic accuracy for all measures. Reasons for this variation in performance are explored below.

The FVMC demonstrated poor sensitivity for the identification of children with LI for ages four through six using a story retell task. Even when the diagnostic criteria were modified to diagnose LI as those that scored in the first percentile (n=3) on the reference measure the SPELT III, the FVMC did not demonstrate an acceptable ability to differentiate LI and TD young school age children ages three years to six years eleven months. However, the descriptive statistics revealed that the children in the TD group did produce statistically significantly more obligatory contexts of the finite verb markers measured by the FVMC across all age groups. Therefore, the data suggested that children with LI do use verb tense morphemes differently than age-matched peers; yet, the FVMC did not demonstrate an ability to capture that difference in language use diagnostically.

The findings were inconsistent with many studies that indicated the FVMC is an acceptable tool for identifying LI in children ages three years to five years ten months (Bedore & Leonard, 1998; Eisenberg & Guo, 2013; Goffman & Leonard, 2000; Guo & Eisenberg, 2013; Souto et al., 2014). Only one study evaluated the clinical utility of FVMC with children ages six years to six years eleven months and found acceptable diagnostic accuracy (Guo & Schneider, 2016). However, it is noteworthy that in that study 36% of six year olds produced FVMC tense morphemes at a level of mastery (over 90%).

Although Reilly et al. (2014) recommended the clinical use of 1.25 SD below the mean to diagnose LI, previous studies have used varied diagnostic criteria. For example, Souto et al.'s study Identify Risk for Specific Language Impairment with Narrow and Global Measures of Grammar (2014) inclusionary criteria for participants with LI included a score in the first percentile rank on the SPELT II. The study found that both narrow and global measures, including the FVMC, had good diagnostic accuracy for four and five year old children. When this present study modified diagnostic criteria to include only those children scoring in the first percentile on the SPELT III, only three children remained in the LI group. Interestingly, three children with LI out of a 141 participant sample did not reflect the clinical prevalence of LI (Tomblin, Records, Buckwalter, Zhang, Smith, & Brien, 1997). The estimated prevalence of specific language impairment in children is 7%, in contrast by diagnosing LI in this study using the first percentile on the SPELT III only 2% of the sample was diagnosed as LI (Tomblin, Records, Buckwalter, Zhang, Smith, & Brien, 1997). Even with the altered diagnostic criteria, the FVMC continued to demonstrate poor sensitivity, as one of the participants with LI performed quite well on the measure, producing 87% of measured verb tense morphemes correctly in 15 total opportunities.

Further inspection of the data revealed interesting patterns at the participant level. Participant 039 had standard score (SS) on the SPELT III within the first percentile, and yet an FVMC score of 87%. The participant was six years five months at the time the language sample was collected, possibly accounting for the proficiency at verb tense morpheme production. This explanation would be consistent with Moyle et al. (2011) who found 45% sensitivity, and 62.5% sensitivity for the FVMC within school age children, ages five years five months to nine years eight months. Moyle (2011) asserted that by the time children reach school age, they have mastered the verb tense morphemes, and therefore the FVMC lost diagnostic power.

An examination of one participant that the FVMC did correctly identify included participant, 047 age 5;6, was within the age range found to be appropriate for the FVMC to be utilized as a diagnostic tool with acceptable diagnostic accuracy (Bedore & Leonard, 1998; Eisenberg & Guo, 2013; Goffman & Leonard, 2000; Guo & Eisenberg, 2013; Souto et al., 2014). Additionally, the participant scored within the first percentile on the reference measure, like those classified as LI in Souto et al.'s (2014) study. Participant 047 earned an FVMC productivity score of 38%. Participant 047F's narrative retell included C-Units such as, "And the frog fell on him face. And the guy wave.... And the frog jumping into the salad," demonstrating a severe deficit in the examined verb tense morphemes at age five years six months. Therefore, the FVMC appears to identify patterns of deficit in finiteness marking. However, not every child with LI demonstrated these deficits, decreasing the diagnostic accuracy and utility of the measure on this study.

Although these examples only explore the language production of two of the thirteen participants with LI, they support Moyle et al.'s hypothesis that by the time children with LI are evaluated at a school-age, they have acquired more verb morphemes. However, this study found that participants with TD produced significantly more verb tense morphemes, and significantly longer utterances than the children with LI. These findings are consistent with prior studies which found children with LI still struggle with morphology in the school years, and that perhaps there is an extended period of time wherein children with LI omit tense markers when compared to TD children (Conti-Ramsden et al., 2001; Rice et al., 1995).

ECU, the other narrow measure examined in this study, did not demonstrate acceptable sensitivity for the identification of children with LI for diagnosing children between the ages of four and six with LI utilizing cutoff scores 1, and 1.25 SD below the mean on the ECU, and when utilizing a reference measure standard score 1.25 SD below the mean on the reference measure. The findings were inconsistent with Guo and Schneider (2016), who found ECU to be a clinically useful tool with acceptable diagnostic accuracy for children ages six and eight. It is noteworthy that ECU was the most sensitive of the measures, and approximated acceptable diagnostic accuracy. ECU demonstrated 76.9% sensitivity, and 96.1% specificity for differentiating children with LI from TD at ages four, five, and six when utilizing cutoff scores that were 1 SD from the mean. It is likely that the measure had the best sensitivity of the measures due to the fine-grained error analysis it provides (Guo & Schneider, 2016). For example, many of the thirteen participants with LI produced pronoun, and possessive errors in addition to verb tense morpheme errors. For example, participant 063, age six years, said, "So he [omitted verb], "Wait, that's my frog." And him were going to bed and laugh." The fine grained analysis provided by the ECU tracked all of the errors produced by this participant to correctly identify the participant as LI. When the reference measure diagnostic criteria were amended to diagnose LI consistently with Souto (2014) by identifying those children who scored in the first percentile (n=3) on the SPELT III, ECU had good sensitivity (100%), and specificity (97%) for differentiating children with and without LI. Although the sample size of this study was small, these results do more closely approximate the results found in Guo and Schneider's (2016) study. It is noteworthy that the wider the disparity in language ability (e.g., creating an LI group that performed in the first percentile on the reference measure, and not one that approximates a clinical sample of children with LI), the better diagnostic accuracy ECU demonstrated on this study.

PGCU did not demonstrate acceptable sensitivity when diagnosing children with LI utilizing cutoff scores 1, and 1.25 SD below the mean on PGCU performance, and when utilizing a reference measure standard score 1.25 SD below the mean below the mean on the reference measure, the SPELT III. The findings were inconsistent with Souto et al. (2014), Eisenberg and Guo (2013), and Guo and Schneider (2016), which found PGCU to be a clinically useful tool with acceptable diagnostic accuracy for children ages four, five, six, and eight. However, when the diagnostic criteria were amended to diagnose LI consistently with Souto (2014) by identifying those children who scored in the first percentile (n=3) on the reference measure the SPELT III, both PGCU had good sensitivity (100%), and specificity (99%) for differentiating young school age children with and without LI. Although the sample size was small, these results do more closely approximate the results found in Souto et al.'s (2014), Eisenberg and Guo's (2013), and Guo and Schneider's (2016) studies. In other words, when applied to participants meeting restrictive diagnostic criteria, the measures sufficiently differentiate the lower performing LI children from TD peers. This is concerning when considering adapting the measure to be applied to a clinical population since diagnosing LI as performing within the first percentile on a standardized battery under identifies language impairment (Tomblin, Records, Buckwalter, Zhang, Smith, & Brien, 1997).

The difference in the diagnostic accuracy of the PGCU is surprising given that some researchers argue that the complex language tasks, such as narrative tasks, should produce more obligatory contexts for difficult morphemes, and prevent children with LI from being able to avoid difficult the forms (Oetting & Horohov, 1997; Thordardottir, 2008). However, Rice and Wexler (1996) examined language productions during picture description and during conversation and did not find any significant differences in language use. One reason may be that the child can avoid the difficult language contexts. For example, Participant 094, age five years ten months, produced approximately three utterances that traditionally retold the pictured story. Conversely, he produced 29 utterances about a tangential topic. For example, "And he said, 'Giddy up horsey.' And he said, 'Yeehaw.' Cop, you go away. You're bad." He later says, "I can't do it. I don't know. I don't know," and abruptly ends his story retell before the actual end of the wordless picture book. In this way, it is possible to propose that certain LI participants have become proficient at avoiding difficult language tasks regardless of elicitation method, limiting the ability for error analysis.

Limitations

One limitation of the present study is that although the sample was predominantly comprised of white (94%), non-Hispanic (90%) children, all language transcripts were analyzed according to the morphosyntactic rules of Standard American English. For a small portion of the sample this pattern of error analysis may have been inappropriate and could have possibly misrepresented those participants' language abilities by marking dialectal differences as errors. Although efforts were made to accept dialectal variations, it is difficult to assess grammaticality without knowing the language environment and cultural context of each participant while analyzing language productions.

In addition to potential dialectal differences playing a role in error analysis, all three supplemental measures may have been limited by the size of the language samples.. The lower bounds, and means of MLU, and obligatory contexts, displayed in Table 3, may be used to gauge the language production in terms of size and opportunity for errors. This is important because measures of grammaticality and finite morphology depend upon production errors (Bedore & Leonard, 1998). Interestingly, all three variables indicated that not all participants produced enough language for adequate analysis. The lower bound for the MLU in morphemes was 2.64, limiting the opportunity for error pattern analysis. The lower bound for the FVMC was zero, and the upper bound was 31.54. Examination of the language analyses of high performing participants with LI exposes how the limited language impacts the limited opportunity for error production. For example, participant 119, age five years eight months, was not classified as LI by FVMC, PGCU, or ECU. At five years eight months, the participant did not produce any obligatory contexts to be analyzed by the FVMC. Although he did produce several ungrammatical utterances, his PGCU was within 1 SD of performance. His developmentally appropriate scores on the supplemental measures are likely due to his low MLU of 3.19, and the low number of total words. For example, he produced the utterance, “I don’t know,” fifteen times in a 25-utterance sample. Therefore, the analyses were inappropriate for analyzing language production when the child’s productions were extremely limited. Future studies would benefit from analyzing larger language samples.

Conclusions and Future Directions

In conclusion, this study did not find FVMC, PGCU, or ECU as acceptable diagnostic tools for children ages four through six. The measures may be more appropriate to supplement information about how the child uses language in different contexts, such as narrative generation or retell. In this way, the error pattern analysis may add information to the language profile of a child after a diagnosis has been made. It is also important to apply the measures cautiously since children with LI may be proficient at avoiding difficult language tasks. Either narrow measure may still provide valuable information about a specific child’s error patterns, which could help create functional therapy targets.

The present study found significant differences in MLU and morpheme use, consistent with the findings of many previous studies (Bedore & Leonard, 1998; Nippold 2009; Nippold, Mansfield, Billow, & Tomblin, 2008). Since many studies have confirmed the significant difference between MLU between LI and TD children, this difference in complexity and production should be explored further by future studies for diagnostic measures that may yield better diagnostic accuracy (Bedore & Leonard, 1998; Nippold 2009; Nippold, Mansfield, Billow, & Tomblin, 2008).

References

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. Journal of Speech, Language, and Hearing Research, 41(5), 1185-1192.

Beyer, T., & Kam, C. L. (2012). First and second graders' interpretation of Standard American English morphology across varieties of English. First Language, 32(3), 365-384.

Cleave, P. L., & Rice, M. L. (1997). An Examination of the Morpheme BE in Children With Specific Language Impairment. *J Speech Lang Hear Res Journal of Speech Language and Hearing Research, 40*(3), 480.

Colozzo, P., Gillam, R. B., Wood, M., Schnell, R., & Johnston, J. R. (2011). Content and form in the narratives of children with specific language impairment. Journal of Speech, Language, and Hearing Research, 54, 1609-1627.

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.

Dollaghan, C. A. (2007). The handbook of evidence-based practice in communication disorders. Baltimore: Brookes.

Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. Journal of Speech, Language, and Hearing Research, 39(3), 643-654.

Ebert, K. D., & Scott, C. M. (2014). Relationships between narrative language samples and norm-referenced test scores in language assessments of school-age children. Language, Speech, and Hearing Services in Schools, 45(4), 337-350.

Eisenberg, S. L., & Guo, L.-Y. (2013). Differentiating children with and without language impairment based on grammaticality. Language, Speech, and Hearing Services in Schools, 44(1), 20-31.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. Journal of Speech, Language, and Hearing Research, 47(6), 1301-1318.

Gillam, R., & Pearson, N. (2004). Test of Narrative Language. Austin, TX: Pro-Ed.

Gladfelter, A., & Leonard, L. B. (2013). Alternative tense and agreement morpheme measures for assessing grammatical deficits during the preschool period. Journal of Speech, Language, and Hearing Research, 56(2), 542-552.

Greenslade, K., Plante, E., & Vance, R. (2009). The diagnostic accuracy and construct validity of the structured photographic expressive language test preschool: Second edition.Language Speech and Hearing Services in Schools, 150-150.

Guo, L.-Y., & Eisenberg, S. (2014). The diagnostic accuracy of two tense measures for identifying 3-year-olds with language impairment. American Journal of Speech-Language Pathology, 23(2), 203-212.

Guo, L., & Schneider, P. (2016). Differentiating School-aged Children with and without Language Impairment Using Tense and Grammaticality Measures from a Narrative Task. *Journal of Speech, Language, and Hearing Research*.

Lee, L. (1974). Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians. Evanston, IL: Northwestern University Press.

Lijmer, J.G., Mol, B.W., Heisterkamp, S., Bonsel, G.J., Prins, M.H., van der Meulen, J.H., & Bossuyt, P.M. (1999). Empirical evidence of design-related bias in studies of diagnostic tests. JAMA: The Journal of The American Medical Association, 1061-1066.

Loban, W. D. (1976). Language development: Kindergarten through grade twelve. Urbana, IL: National Council of Teachers of English.

Mayer, Mercer (1974). *Frog goes to dinner*. New York: Dial Books for Young Readers.

Moyle, M., Karasinski, C., Ellis Weiser, S., & Gorman, B. (2011). Grammatical morphology in school-age children with and without language impairment: A discriminant function analysis. Language, Speech, and Hearing Services in Schools, 42, 550-560.

Nippold, M. (2009). School-age children talk about chess: Does knowledge drive syntactic complexity? Journal of Speech, Language, and Hearing Research, 52, 856-871.

Nippold, M., Mansfield, T., Billow, J., & Tomblin, J. B. (2008). Expository discourse in adolescents with language impairments: Examining syntactic development. American Journal of Speech-Language Pathology, 17, 356-366.

Oetting, J. B., & Horohov, J. E. (1997). Past-tense marking by children with and without specific language impairment. Journal ofSpeech, Language, and Hearing Research, 40, 62-74.

Paul, R., & Norbury, C. F. (2012). Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating (4th ed.). St. Louis, MO: Elsevier.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. Language, Speech, and Hearing Services in Schools, 25(1), 15-24.

Plante, E., & Vance, R. (1995). Diagnostic accuracy of two tests of preschool language.

American Journal of Speech-Language Pathology, 4, 70–76.

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of Clinical Markers of Specific Language Impairment in Adults. *J Speech Lang Hear Res Journal of Speech Language and Hearing Research, 53*(2), 414.

Reilly, S., Tomblin, B., Law, J., Mckean, C., Mensah, F. K., Morgan, A.,…Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language & Communication Disorders INT J LANG COMMUN DISORD, 49*(4), 416-451.

Rice, M., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. Journal of Speech, Language, and Hearing Research, 39(6), 1239-1257.

Rice, M. L., & Wexler, K. (2001). Rice/ Wexler Test of Early Grammatical Impairment. San Antonio, TX: The Psychological Corporation.

Rice, M., Wexler, K., & Cleave, P. (1995). Specific language impairment as a period of extended optional infinitive. Journal of Speech and Hearing Research, 38, 850-863.

Rutjes, A., Reitsma J.B., Vandenbroucke, J.P., Glas, A.S., Bossuyt, P. (2005). Case-control and two-gate designs in diagnostic accuracy studies. Clinical Chemistry, 51, 1335-1341.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. Journal of Speech, Language, and Hearing Research, 43(2), 324-339.

Souto, S. M., Leonard, L. B., & Deevy, P. (2014). Identifying risk for specific language impairment with narrow and global measures of grammar. Clinical Linguistics & Phonetics, 28(10), 741-756.

Snowling, M. J., Duff, F. J., Nash, H. M., & Hulme, C. (2015). Language profiles and literacy outcomes of children with resolving, emerging, or persisting language impairments. *Journal of Child Psychology and Psychiatry, 57*(12), 1360-1369.

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? Language, Speech, and Hearing Services in Schools, 37(1), 61-72.

Thordardottir, E. (2008). Language-specific effects of task demands on the manifestation of specific language impairment: A comparison of English and Icelandic. Journal of Speech, Language, and Hearing Research, 51(4), 922-937. doi: 10.1044/1092-4388(2008/068)

Tomblin, J. B., Records, N. L., 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*(6), 1245.

Wiig, E. H., & Secord, W. (1989). Test of Language CompetenceV Expanded Edition. San Antonio, TX: The Psychological Corporation.

Wiig, E. H., Secord, W. A., & Semel, E. (1992). Clinical Evaluation of Language Fundamentals -Preschool. San Antonio, TX: The Psychological Corporation

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2002). Preschool Language Scales—Fourth Edition. San Antonio, TX: The Psychological Corporation.

Table 1. Word and Utterance Level Codes for SALT

Generated code	*Meaning*	*Example*
[T]	tense error (third person singular -s, past tense -ed)	And he went to bed and laugh[T] and laugh[T].
[PR]	pronoun error (subject pronouns, object pronouns, reflexive pronouns, possessive pronouns, and possessive determiners)	Him[PR] pat him[PR] dog.
[-PR]	omission pronoun	*It[-PR] looks like one is near the water.
[GM]	Grammatical morpheme error (plural -s, prepositions, and present and past participles)	We played in[GM] the beach.
[-GM]	Grammatical morpheme omission	The frog was looking out *of[-GM] his pocket.
[BE]	auxiliary or copula be error	Why's[BE] them in there playing?
[-BE]	omission auxiliary or copula be	The boy *is[-BE] taking the frog out.
[A]	article error	I want a[A] chicken nuggets.
[-A]	omission article	Go on *article[-A] slide.
[S]	semantic error	He's blobby blobby[S].
[-SUB]	omission subject	And *subject[-SUBJ] landed into it.
[-OBJ]	omission object (argument structure errors as omissions of objects in any contexts surrounding verbs)	I falled on *object[-OBJ]..
[OVE]	other verb error including overgeneralization	Last time he come[OVE] in the gym.
[EW]	other errors including syntactic or semantic errors that did not fit the any of the aforementioned prescribed categories (e.g., article errors, omission possessives)	And then the cat jumped and got the frog right he saw him landed on baby/z lap[EW].
[G]	The utterance is considered grammatical.	It jumped out of the lettuce [G].
[UG]	The utterance is considered ungrammatical.	When they ate the restaurant the froggy peeked out [UG].

Table 2. Group Differences by Age Group

		df	F	Sig.
FVMC	Between Groups	2	1.546	.217
	Within Groups	138		
	Total	140		
PGCU	Between Groups	2	1.59	.208
	Within Groups	138		
	Total	140		
ECU	Between Groups	2	1.84	.163
	Within Groups	138		
	Total	140		
MLUm	Between Groups	2	10.66	.000
	Within Groups	138		
	Total	140		
Obligatory Contexts	Between Groups	2		
	Within Groups	138	9.93	.000
	Total	140		

Table 3. Descriptive Statistics by Age Group

	4-Year-Olds (n=36)	5-Year-Olds (n=69)	6-Year-Olds (n=36)
	Mean (SD) Range	Mean (SD) Range	Mean (SD) Range
FVMC	92.39 (10.51) 60-100	93.32 (14.35) 0-100	97.01 (7.86) 55.17-100
PGCU	78.66 (17.38) 12.5-96.77	82.38 (17.37) 33.33-100	84.49 (13.96) 37.93-100
ECU	.24 (.24) .03-1.19	.20 (.15) .00-.77	.16 (.17) .00-.85
MLUm	31.81 (13.00) 3.35-9.68	6.45 (1.55) 2.64-12.21	8.25 (1.59) 3.92-11.28
Obligatory	19.22 (7.99) 5.00-37.00	24.52 (9.33) 1.00-64.00	28.52 (8.91) 1.00-64.00

Table 4. Post Hoc Analysis by Age Group

		MlUm		*Obligatory Contexts*
		Sig.		Sig.
4-Year-Olds (n=36)	5	0.002	5	0.012
	6	0.000	6	0.000
5-Year-Olds (n=69)	4	0.002	4	0.012
	6	0.197	6	0.077
6-Year-Olds (n=36)	4	0.000	4	0.000
	5	0.197	5	0.077

Table 5. Descriptive Statistics by Language Status

	TD (n=128)	LI (n=13)		
	Mean (SD)	Mean (SD)	F	Sig.
FVMC	96.09 (5.98)	73.39 (28.72)	80.85	0.00
PGCU	84.10 (10.74)	59.97 (23.69)	27.68	0.00
ECU	0.17 (0.14)	0.48 (0.31)	23.67	0.00
MLUm	7.64 (9.82)	5.87 (1.90)		0.00
Obligatory Contexts	24.92 (9.30)	17.00 (8.89)		0.00

Table 6. FVMC Sensitivity and Specificity

FVMC					
Cut-off Scores	n	Sensitivity %	Specificity %	LR+	LR-
1 SD Cut-off at 4=81.88, at 5=78.97 and at 6=89.14	LI=7 TD=134	53.80	99.20	7.54	.47
1.25 SD Cut-off at 4=79.26, at 5=75.29 and at 6= 87.18	LI=6 TD=135	46.15	96.18	12.09	.56
2 SD Cut-off at 4=71.40, at 5=64.62 and at 6=81.28	LI=2 TD=139	66.66	97.16	23.5	.24

Table 7. PGCU Sensitivity and Specificity

PGCU					
Cut-off Scores	n	Sensitivity %	Specificity %	LR+	LR-
1 SD Cut-off at 4=61.28, at 5=70.16 and at 6=70.53	LI=8 TD=133	61.54	94.66	11.52	.41
1.25 SD Cut-off at 4=56.93, at 5=67.10 and at 6= 67.04	LI=7 TD=134	53.85	94.61	10	.49
2 SD Cut-off at 4=43.92, at 5=57.93 and at 6=56.57	LI=3 TD=138	100	99.29	141	0

Table 8. ECU Sensitivity and Specificity

ECU					
Cut-off Scores	n	Sensitivity %	Specificity %	LR+	LR-
1 SD Cut-off at 4=0.48, at 5=0.35 and at 6=0.33	LI=10 TD=131	76.92	96.18	20.15	.02
1.25 SD Cut-off at 4=0.54, at 5=0.38 and at 6= 0.37	LI=7 TD=134	53.85	96.18	14.11	.48
2 SD Cut-off at 4=0.72, at 5=0.50 and at 6=0.50	LI=3 TD=138	100	97.16	35.25	0