Tense Marking in the General Kindergarten Population: Is there Evidence of Bimodal Distribution?

By

Brian Kenneth Weiler

Dissertation
Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY
in

Hearing and Speech Sciences
May, 2016
Nashville, Tennessee

Approved:

Professor C. Melanie Schuele

Professor Stephen M. Camarata

Professor James W. Bodfish

Professor Megan M. Saylor

To Courtney, Quinn, Emma, and Lillian

ACKNOWLEDGEMENTS

This study was made possible through the support of a US Department of Education Preparation of Leadership Personnel grant (H325D080075; PI: Schuele), an American Speech-Language-Hearing Foundation Student Research Grant in Early Childhood Language Development (PI: Weiler), and a Vanderbilt Institute for Clinical and Translational Research Voucher.

Completion of this study would not have been possible without the guidance and mentorship of my advisor Dr. C. Melanie Schuele. Her advice to "keep putting one foot in front of the other" kept me on track when the path forward seemed daunting. Through her tireless work ethic and her consideration of others above herself, Dr. Schuele is a true leader by example.

I am further appreciative of the helpful advice and constructive feedback provided by my other committee members, Drs. Stephen Camarata, James Bodfish, and Megan Saylor. I thank Dr. Warren Lambert for assistance with the statistical aspects of this study. I am grateful to the members of the Vanderbilt Child Language and Literacy Lab — Jacob Feldman, Hannah Krimm, Magdalene Jacobs, Sylvia Liang, and Marley Kern — for their contributions to data collection, test scoring, and reliability coding.

I owe a debt of gratitude to the teachers, administrators and staff who generously invited me into their wonderful elementary schools to complete this study. Finally, I thank the kindergarten children and their parents for participating in this research; it was a privilege and delight working with them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

INTRODUCTION

Children with specific language impairment (SLI) comprise 7.4% of the kindergarten population (Tomblin et al., 1997). These children, who demonstrate language difficulties in the face of otherwise normal development (Leonard, 2014), are notoriously under-identified and, by extension, under-served. In Tomblin and colleagues' classic epidemiological study of 7,218 kindergarteners, parents of only 29% of the 216 children who qualified as SLI according to the research criteria reported that their child had been referred previously for clinical services due to concern related to speech and/or language development (1997). Among those kindergarteners for whom only language was impaired, only 9% had ever received intervention services (Zhang & Tomblin, 2000). When speech and language were impaired, the rate of intervention receipt was 41%. In essence, the marginal likelihood for therapy referral was limited primarily to those students with SLI and poor speech articulation. Given that the co-occurrence of speech sound disorder and SLI in a subsample of this same six-year-old study population (n = 1328) was only 5-8% (Shriberg, Tomblin, & McSweeny, 1999), we are left with the sobering message that kindergarteners with SLI, the majority of whom have unremarkable speech articulation, fall way below the radar.

This finding might not be troublesome if kindergarten language impairment were a fleeting phenomenon that later resolves. Quite to the contrary, these children, followed longitudinally, continue to lag behind peers with typical language throughout adolescence not only in academic tasks such as math and reading, but also in the areas of social participation and

self-esteem (Tomblin, 2008). Compromised academic, behavioral, psychosocial, and vocational outcomes have been documented with other samples of children with language impairment followed longitudinally (Beitchman, 2001; Conti-Ramsden, Durkin, Simkin & Knox, 200 9; Johnson et al., 1999; Law, Rush, Schoon, & Parsons, 2009; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998).

Clearly, if the academic, behavioral, and socio-emotional disadvantages conferred to children with language impairment are to be minimized through the receipt of services (via special education), then these children first have to be identified. The incongruence between research prevalence and prior confirmation of language impairment reported by Tomblin et al. (1997) is strongly suggestive of a problem of identification. This problem was addressed by an expert panel at the National Institutes of Health (NIH) who called for continued research to identify clinical markers of SLI (Tager-Flusberg & Cooper, 1999). A clinical marker can be considered a linguistic form that is characteristic of and especially sensitive to the diagnosis of a language impairment. The NIH panel noted that a "composite reflecting children's degree of use of several finite verb-related morphemes in obligatory contexts seems to hold considerable promise, at least for English, as a measure that distinguishes children with SLI from their typically developing peers" (p. 1276).

Tense marking and the construct of finiteness

The term finiteness relates to a small set of verb-related morphemes that, in English, carry the *tense* and *agreemen*t features that are obligatory in the matrix clause (Rice, 2004). The morphemes that mark finiteness can be free-standing as is the case with *BE* copula and auxiliary (e.g., *Emma <u>is</u> happy; Quinn and Lillian <u>are</u> playing*) and irregular past tense (e.g., *Courtney <u>ran</u>*). Other finiteness-marking morphemes, such as regular past tense —*ed* (e.g., *Quinn jump<u>ed</u>*)

and third-person singular present tense—*s, es* (e.g., *Emma laughs*) are affixed to lexical verbs. Some finiteness markers, such as past tense (PT), carry only the tense feature of the clause. Others, such as third-person singular present tense (3S), carry the *tense* and subject-verb *agreement* features of the clause (compare *Every day Lillian laughs* to *Every day they laugh_*). The group of finiteness-marking morphemes collectively is considered part of a grammatical computational system related to the acquisition of grammatical well-formedness during the preschool and early school-age years (Rice, 2004). For the purposes of brevity and consistency with common clinical and research nomenclature, the term *finiteness marking/markers* will be referred heretofore as *tense marking/markers*.

Rates of obligatory tense marker omissions (e.g., *She __ running; Yesterday he play__*) reliably distinguish children aged 3-8 years with SLI from same-age peers with typical language (TL). Over a dozen studies have reported noticeable separation of performance (median Z-score of -4.59) between the two groups (for a review see Ash & Redmond, 2014). The utility of tense marking to meaningfully separate SLI and TL groups has proven stable longitudinally (Rice, Wexler, & Hershberger, 1998) and across data collection methods, including conversational samples and sentence elicitation tasks (e.g., Krok & Leonard, 2015; Rice & Wexler, 1996) as well as sentence recall tasks (e.g., Abel, Rice, & Bontempo, 2015; Hoover, Storkel, & Rice, 2012).

The bimodality hypothesis of kindergarten tense marking

Instead of being distributed normally (i.e., a bell curve distribution), tense-marking proficiency at the point of school entry (that is, kindergarten) has been hypothesized to follow a bimodal distribution; children with SLI cluster at the lower end of the distribution whereas children with typical language cluster toward the upper end as they are approximating "adult

grammar" (Bishop, 2004; Rice, 2000). From an identification standpoint, a clinical marker distributed bimodally considerably reduces the arbitrariness of the criterion with which one determines "affectedness" (Spaulding, Plante, & Farinella, 2006; Spaulding, Swartwout Szulga, & Figueroa, 2012). Common diagnostic and service eligibility standards often dictate, for example, that scores 1.5 standard deviations (SD) below a normative mean signify the presence a delay/disorder (e.g., Colorado Department of Education, 2010; Tennessee Department of Education, 2009). Such statistical cutoffs, however, run a risk of arbitrarily dichotomizing a continuous metric when applied to a normally-distributed skill because some children falling above or below the threshold will meet inclusionary or exclusionary criteria otherwise (Bishop, 2014; Tomblin et al., 1997). Is there really a meaningful difference in the likelihood of the presence of functional impairment between, for example, a score that is 1.4 SD below the mean compared to a score 1.6 SD below the mean? Instead, if kindergarten tense marking is indeed a bimodally-distributed skill, then it should allow for easier and more valid identification because a clear boundary would separate the performance of children with SLI from their TL peers.

Theoretical considerations and corroborating evidence

To be clear, the debate over whether children with SLI represent a qualitatively distinct subgroup in the population is unresolved. Tomblin and Zhang (1999) captured the essence of this debate as centering on the question: "Are children with SLI a different group of language learners who have a distinctive form of linguistic behavior . . . [o]r are these children most likely to be the tail end of the distribution of normal language learners?" (p. 362). Leonard (1991) took a strong stance for the latter position, arguing that many children with SLI "may be different solely because they fall on the very low end of the normal distribution in ability" (p. 68).

The theoretical orientation from which the basis of SLI is viewed appears to largely drive the debate over how best to characterize this diagnosis. As summarized by Kamhi (1998), the view of SLI as characterized by differences in processing capacity (e.g., Leonard, 1994) is consistent with a continuum model. Within this model, "the exaggerated profiles" of children with SLI "are a natural outcome of a continuum of language abilities" (Leonard, 2014, p. 4). Accordingly, such children are seen as representing the lower end of a continuum. Such a "differences in degree but not kind" stance is supported by several studies that have failed to find separable and distinct diagnostic categories (e.g., SLI, TL) based on taxometric, latent class analyses of test score distributions. Dollaghan (2004) did not find evidence of a distinguishable diagnostic classification of SLI in four-year-olds using scores from measures of receptive vocabulary and mean length of utterance. In addition, Dollaghan (2011) failed to find latent class evidence for an SLI category in six-year-old children using scores from measures of expressive lexical diversity (number of different words) and phonological working memory (nonword repetition task). She qualified this finding by noting that "additional analyses on other diagnostic measures of SLI clearly are necessary before strong conclusions about latent structure can be drawn" (p. 1369). Dollaghan specifically noted a measure of grammatical tense marking, the Rice/Wexler Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001), as an additional promising candidate for latent class analysis. Leonard (2004), too, conceded that "the language measures in these [taxometric] studies have not yet focused on grammatical computation" (p. 4).

The presence of an underlying grammatical deficit characterizing SLI might therefore offer support of a cluster model over a continuum model. Under a grammatical deficit hypothesis, children with SLI would represent a qualitatively different cluster from the general

population when grammatical skill is assessed (Kamhi, 1998). The most compelling evidence to date for this model comes from a study conducted by Rice and Wexler (1996) and further reported in Rice (1998). Similarly-sized groups of clinically-identified children with SLI and age-matched peers with TL were compared on a composite measure of tense-marking accuracy with finite morphemes (e.g., *BE* forms, PT, 3S) across elicitation probes and conversational samples. Rice and Wexler reported that 36 of the 37 five-year-olds in their SLI group marked tense in obligatory contexts with less than 60% accuracy whereas all of the 45 five-year-olds in their normal language control (NC) group marked tense with approximately 80% or greater accuracy (see Figure 1).

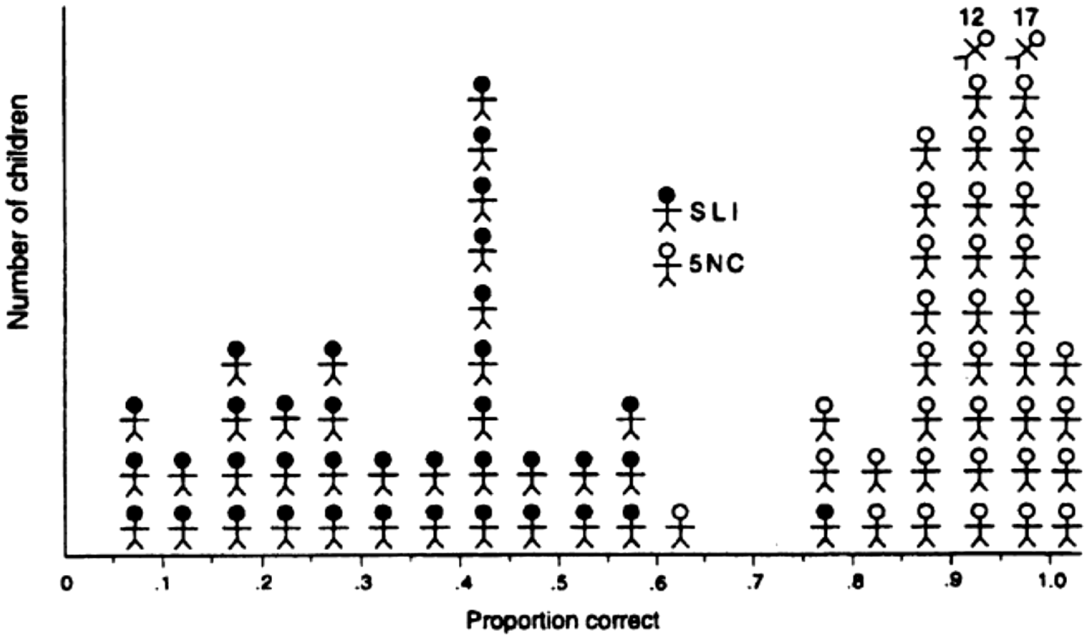

Figure 1. Distribution of individual children's performance on a composite tense marking score: SLI and age controls (Rice, 1998). Copyright © 1998 American Speech-Language-Hearing Association. Reprinted with permission. SLI = specific language impairment; 5NC = five-year-old normal language controls.

Cluster separation is appreciated by the practically non-overlapping "buffer zone" of at least 10 percentage points (i.e., 65-75%) between the two group distributions illustrated in Figure 1. The degree of non-overlap between the two distributions, calculated by converting the Cohen's *d* effect size to a U measure (Cohen, 1988), ranges from 87 to 95% per individual morpheme. That each individual tense marking morpheme from the composite reliably differentiated children in the SLI group from age-peers was taken to indicate that 3S and PT ". . . are not likely to be isolated surface phenomena. Instead, these morphemes serve to mark [tense], as do *BE* and *DO*, and this [tense]-marking feature constitutes a clinical marker" (Rice & Wexler, 1996, p. 1251). From an identification standpoint, these results suggest that the likelihood of false positives and false negatives resulting from the use of a tense-marking composite for identification of SLI should be minimal.

From theory to practice

Indeed, of the psychometric properties for 43 standardized child language assessment measures reviewed by Spaulding and colleagues, (2006), a test of tense marking, the TEGI (Rice & Wexler, 2001) was one of only five assessments for which acceptable levels of sensitivity and specificity ($\geq$ .80; Plante & Vance, 1994) were reported. Of concern, only two of the five standardized assessments meeting acceptable psychometric standards for discriminant validity — the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4; Semel, Wig, & Secord, 2003) and the Preschool Language Scale, Fourth Edition (PLS-4; Zimmerman, Steiner, & Pond, 2002) — are commonly used by speech-language pathologists (SLPs) when assessing children aged 5 to 9 with suspected SLI (Betz, Eickhoff, & Sullivan, 2013). SLPs also commonly use tests of single-word vocabulary, for example the Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007) and the Expressive One-Word Picture

Vocabulary Test, Third Edition (EOWPVT-3; Brownell, 2000), as measures to identify SLI. Because 50% of SLPs working with school-aged children rated standardized tests as the most important assessment measures in their diagnostic protocol (Eickhoff, Betz, & Ristow, 2010), evidence-based practice should guide test selection such that selected diagnostic measures have minimal distributional overlap between typical and language-impaired populations.

Tests of single-word vocabulary, however, appear to tap a linguistic domain (e.g., lexical labeling) with a high degree of overlap between typical-language and language-impaired distributions. This trend can be appreciated by the age-aggregated mean standard deviation standard score difference for typical and language-impaired samples reported in the Peabody Picture Vocabulary Test, Third Edition manual (PPVT-3, Dunn & Dunn, 1996). As reported by Spaulding et al. (2006), a mean standard deviation difference of 0.55 is equivalent to a 20% non-overlap between normal and language-impaired distributions (see Figure 2). With a quick glance at the very minimally overlapping distributions in the aforementioned TEGI data (Figure 1, Rice, 1998), it becomes clear that the clinical marker potential of assessing tense marking for SLI identification, at least around the time of school entry, appears far more psychometrically promising than other more commonly chosen assessment measures.

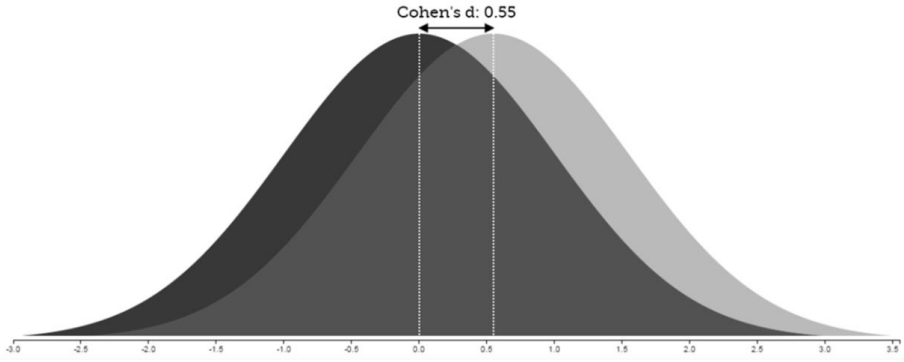

Figure 2. Twenty percent distributional non-overlap associated with age-aggregated mean 0.55 *SD* difference reported in Spaulding et al. (2006) between language-impaired and normal language normative group performance on the Peabody Picture Vocabulary Test, Third Edition (PPVT-3; Dunn & Dunn, 1996).

SLI and matched group designs – Potential limitations and biases

Although theoretical debates abound regarding the potential cognitive-linguistic mechanism underlying poor control of tense-marking morphemes among children with SLI, there is general agreement that such debates should not preclude the clinical use of measures such as the TEGI (Leonard, Miller, & Gerber, 1999). Instead, Leonard and colleagues (1999) recommended pursuing "parallel programs of research, one aimed at evaluating and refining finite verb morphology as a clinical marker, the other at discovering precisely why this area of language should be so prone to difficulty in children with SLI acquiring English" (p. 688). The purpose of this dissertation was to pursue the former recommendation. Specifically, this project endeavored to address a recent challenge to the research methods used in studies that have indicated tense marking as a clinical marker for SLI. In a review of the existing evidence for the inclusionary criteria of SLI, Reilly et al. (2014) noted that claims of high diagnostic accuracy using tense verb morphology come from matched-group designs comprised of an SLI group compared to a typically-developing control group. Reilly and colleagues argued that such a matched-group approach is problematic for evaluating the diagnostic utility of a measure. They noted that:

> These metrics are not simply a function of the reliability of the diagnostic tool but also the prevalence of the disorder in the population being tested. Methodologies that include 30-50% of children with SLI in their samples (i.e., matched group designs) artificially inflate the sensitivity of any test and do not represent a tool's functioning in a population sample, wherein the prevalence would be approximately 7% (Tomblin et al., 1997). (2014, p. 426)

Given the aforementioned report of only 29% clinical identification of SLI in the kindergarten population (Tomblin et al., 1997), it quite possibly the case that the LI samples in matched-group designs using clinically-ascertained samples are comprised primarily of children with more severe deficits rather than children with less severe deficits. Arguably, the children

with more severe deficits are potentially more likely to have been identified for services by kindergarten. Such a clinically-sampled group is inherently biased when compared to the whole population because the group may represent a "phenotypically enriched" sample (Mueller, 2012). Children with more mild-moderate SLI, on the other hand, may very well be under-represented in matched group design studies that use clinical identification or clinical referral for recruiting participants (Spaulding et al., 2006). On the other side of the sampling equation, bias may be introduced when comparison groups of unaffected children demonstrate above-average abilities that do not reflect the population mean. Watkins and Johnson (2004), in a review of research principles in studies of language and stuttering in young children, reported that the control group in many such studies performed .5 *SD* to 2 *SD* above the population mean on measures of language skills. The potential for control group sampling bias also is reflected in a meta-analysis that reported an average .7 *SD* nonverbal intellectual quotient (IQ) advantage for age-matched peers when compared to children with SLI (Gallinat & Spaulding, 2014). At both ends of the participant spectrum, therefore, the cumulative effect of unintended sampling bias in matched-group research designs may artificially exaggerate group differences that otherwise might be moderated in the general population.

The validity of findings from matched-design studies pointing toward a bimodal distribution of tense marking proficiency can be called into question on the grounds that any "clear separation" boundary reported between affected and unaffected groups might have been blurred had the range of skill been designed to vary as it does in the general population. Law, Reilly, and Snow (2013) commented on the distinct likelihood of such a scenario:

> It is highly unlikely that those presenting to clinics will be representative of the population with communication impairments as a whole, and this means that the results of clinical research studies should only be generalized back to that sample (i.e., the clinical sample and not the whole population from which they are derived). (p. 489)

Instead, the most compelling evidence for a measure's diagnostic accuracy comes from large samples representing the full variation and heterogeneity of the target condition in the population (Dollaghan, 2004; Sackett, Straus, Richardson, Rosenberg, & Haynes, 2000; Tager-Flusberg & Cooper, 1998).

Single-gate versus two-gate designs

Taking a cue from recommended best practices in clinical epidemiology, a sound methodological approach to capturing a representative spectrum of the target condition — and to minimize sampling and selection bias — is the use of a *single-gate* recruitment design (Leeflang, Bossuyt, & Irwig, 2009). Single-gate designs — which are sometimes referred to as 'cohort type accuracy studies' (Bossuyt & Leefland, 2008) — involve a study sample:

> . . . intended to be unselected, comprising a large, broad, and representative swath of individuals that will presumably include some with and some without the disorder. (Dollaghan & Horner, 2011, p. 1078)

The alternative approach for recruiting participants for a diagnostic accuracy study is a *two-gate* design. In a two-gate design, participants are selected and assigned to an 'affected' or a 'control' group a priori based on the known presence or absence of the target condition. Clinical referral is a primary mechanism for assignment to an affected group. Although common in diagnostic studies, two-gate designs can introduce a serious source of spectrum bias if participants from the respective groups to be compared come from different populations or are not fully representative of the target condition (Dollaghan & Horner, 2011; Leeflang et al., 2009; Reitsma et al., 2009). If the affected group in a two-gate study is comprised of clinically-referred individuals who exhibit a greater severity of symptoms than would otherwise be observed in a single-gate study, then the diagnostic accuracy of the index measure runs the risk of being inflated when compared to its accuracy in a single-gate study (Battaglia et al., 2001; Dollaghan & Horner, 2011).

Two recent meta-analyses of studies addressing the diagnostic accuracy of SLI using measures of tense marking highlight an overwhelming reliance on two-gate designs (Leonard & Krok, 2015; Pawlowska, 2014). Of the 23 unique studies reported across the two meta-analyses, only two utilized single-gate designs.[1]

In the first single-gate study, Poll, Betz, and Miller (2010) examined grammaticality judgements of tense marker omissions in complex sentences. Participants were 31 adults aged 18 to 25 attending a vocational post-secondary school. This school was judged by Poll et al. to be more likely to enroll persons with learning disabilities than a four-year college. In keeping with Tomblin et al.'s (1997) finding that less than one-third of kindergarten children with SLI previously had been diagnosed, Poll and colleagues did not require a history of language impairment for adult SLI classification. Instead, the researchers classified participants based on their own testing criteria using the Test of Adolescent and Adult Language, Third Edition (TOAL-3; Hammill, Brown, Larsen, & Wiederholt, 1994) and the Peabody Picture Vocabulary Test, Revised (PPVT–R; Dunn & Dunn, 1981). Not surprisingly given Tomblin's findings, only one of the 13 participants meeting Poll et al.'s SLI criteria reported having received language therapy. Eight others reported past academic difficulties, primarily with reading. Diagnostic accuracy results demonstrated that the grammaticality judgement task had a high level of specificity (.94) but not sensitivity (.54). Because Poll et al. (2010) is the only known study of tense proficiency in adults, it is not possible to evaluate its outcomes relative to a comparable two-gate study.

[1] Another study not included in these meta-analyses (Redmond, Ash, & Hogan, 2015) utilized a pooled study sample of 7- to 9-year-old children with SLI that was clinically ascertained (n = 8) as well as sourced from school-wide language screenings and follow-up confirmatory testing (n = 11). A measure of tense marking (TEGI) was administered, but because the reported results are not disaggregated by sampling method, it is not possible to determine only the performance of the single-gated sample from the screenings.

The second single-gate study concerning tense marking in SLI allows for a comparison of findings relative to similar single-gate studies. Rice, Tomblin, Hoffman, Richman, and Marquis (2004) acknowledged the need to test the generalizability of previous findings from clinically-ascertained samples to a larger population-based sample of unidentified children with SLI. They compared the tense-marking accuracies from a subgroup of kindergarten children with SLI (n = 187) from Tomblin et al.'s (1997) epidemiologically drawn sample to age-matched controls (n = 141) from the same cohort. Statistically significant group differences were found. However, the degree of non-overlap between the SLI and control groups on the composite tense measure (3S and PT) was considerably less robust (25%; d = .65) than that reported for Rice and Wexler's (1996) clinically-ascertained SLI sample (87-95%; ds = 3.19-3.45).

There are at least two factors that might explain the disparity in between-group magnitudes of tense-marking difference between the single-gate and two-gate studies conducted by Rice and colleagues. First, the strength of using tense marking as a clinical marker may be diminished when the range of tense-marking deficits in SLI (i.e., mild to severe) is more fully represented in the participant sample. In other words, the hypothesized bimodality of distribution for this skill may be less obvious when assessed in an epidemiologically-ascertained single-gate, as compared to a clinically-ascertained two-gate, sample. Second, some children who comprised Tomblin et al.'s kindergarten SLI group met the research diagnostic criteria because of poor vocabulary and/or narrative but otherwise intact grammar skills. The language criteria for SLI inclusion in Tomblin et al.'s (1997) epidemiological study was performance at least 1.25 *SDs* below the mean on at least two of the five Composites (see Figure 3) derived from subtests of the Test of Language Development-Primary, Second Edition (TOLD:P–2; Newcomer

& Hammill, 1988) and a narrative story task involving comprehension and production (Culatta, Page, & Ellis, 1983).

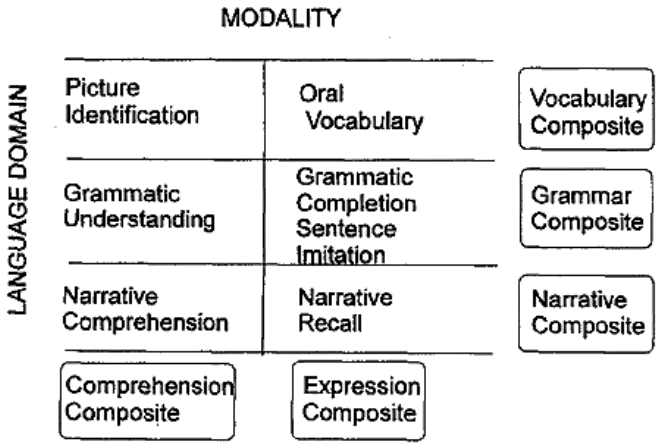

Figure 3. The areas of language measured in the epidemiological SLI study and the composite scores derived from these measures (Tomblin, Records, & Zhang, 1996). Copyright © 1996 American Speech-Language-Hearing-Association. Reprinted with permission.

Collapsing across the Comprehension and Expression modalities, Tomblin and Zhang (1999) graphed the percent of children in the SLI group according to which of the three language domain(s), Vocabulary, Grammar, Narrative, they failed (i.e., scored –1.25 *SD*s). Approximately 35% of the children in the SLI group met criteria because of poor performance in only the Vocabulary and/or Narrative domain(s) of language. In other words, at least one-third of the SLI group from the Tomblin epidemiological study failed to demonstrate general grammatical weakness on testing. The utilization of this SLI subgroup in Rice et al.'s (2004) retrospective study of tense marking may have served to constrict the performance boundary between the SLI and age-matched groups on the grammatical tense measures. By contrast, all children in the SLI group from Rice and Wexler's (1996) two-gate study met inclusionary criteria for general weakness in grammatical development (i.e., mean length of utterance at least 1 *SD* below the age norms of Leadholm & Miller, 1993).

The Present Study

Regardless of the factors at play in the differences between single-gate and two-gate study design findings of kindergarten tense-marking distributions, there is a clear research need for further empirical testing. Rice et al.'s (2004) single-gate design findings replicated previous findings from clinically-ascertained two-gate designs insofar as generalizing evidence of group differences in kindergarten tense marking proficiency to a "broader group of unidentified children affected with SLI" (p. 828). Evidence of the generalization of a bimodal distribution of tense marking to a general kindergarten population, on the other hand, remains to be established. Single-gate studies are not immune to spectrum bias. The (in)accuracy of a measure in one site or sample may not generalize to another sample (Dollaghan & Horner, 2011). The children from Rice et al.'s (2004) sample came from a larger epidemiologically drawn cohort residing in a limited Midwest region of the United States (Tomblin et al., 1997). Tomblin et al. (1997) noted that "it is not possible to claim that these children are fully representative of the U.S. population" (p. 1257).

To further refine an understanding of how tense-marking proficiency is represented in the kindergarten population — and, by extension, to further test the validity of a hypothesized bimodal distribution — requires moving beyond a group design approach. "The crucial need," wrote Dollaghan (2004), "is for strong empirical tests of indicators that are proposed to be diagnostic; neither theoretical preferences nor group comparison studies provide adequate evidence in this regard" (p. 467). Taxometric methods offer a way to empirically evaluate this identified need. Taxometric methods allow for an examination of whether the fundamental latent structure of a given construct is categorical (taxonic) or continuous (dimensional) in nature (Ruscio & Ruscio, 2004). As outlined in Ruscio and Ruscio (2004), "three broad families of

analytic techniques traditionally have been used to test for taxonic boundaries: cluster analysis, finite mixture modeling, and latent class analysis" (p. 155). Cluster analysis was employed in the present study as a technique to evaluate the extent to which kindergarten children indeed "cluster" within a bimodal distribution when the construct of finiteness is assessed via tense-marking accuracies.

In the present study, tense marking was assessed with the TEGI Screening Test in a population-based sample of kindergarteners for whom no a priori classification or grouping criteria was applied. If the bimodality hypothesis was confirmed in the data, then it was expected that an empirically-derived two-cluster distribution of TEGI Screening Test Scores would result. Moreover, given prior epidemiological evidence that the TOLD:P-2 subtest of Grammatical Completion (for both tense morphemes and non-tense morphemes like plural and possessive "-s") posed the greatest difficulty for kindergarteners meeting the research definition of SLI (Tomblin & Zhang, 1999), it was hypothesized that membership in the low-performing distribution (cluster) would broadly hover around Tomblin et al.'s (1997) prevalence rate for language impairment. If, on the other hand, analyses revealed either a single continuous structure or a multiple cluster structure with no obvious or meaningful boundaries, then the validity of treating tense as a bimodally-distributed skill — and hence the diagnostic utility of assessing this skill in the general population, for example as a kindergarten-wide screener for SLI — must be called into question.

The present study explored whether evidence for a bimodal distribution of tense marking indeed exists in the general kindergarten population. Two research questions were addressed:

1. *Do composite tense-marking scores collected from a population-based sample of kindergarten children within a single school district distribute non-normally?*
2. *Do composite tense-marking scores from the TEGI Screening Test suggest the existence of a latent class of children with language impairment who cluster together, apart from a separate latent class of children with typical language?*

CHAPTER II

METHOD

Targeted Kindergarten Population

In the present study, cluster analysis, a conventional approach for testing categorical (or taxonic) boundaries, was run on data collected as part of an ongoing grant-funded study of the grammatical skills of kindergarten children (Weiler, 2014). Specifically, this analysis focused on data collected in the fall of the 2014-15 school year within one public school district in middle Tennessee (TN). This school district is situated in a county that, according to 2010 U.S. Census data, is overwhelming rural. Nearly all of the geographic land area of the county (99.3%) is rural as opposed to urban. The majority, or 82.5%, of the county's 2010 total population of 18,538 is represented by a rural population (U.S. Bureau of the Census, 2010). In the epidemiological study (Tomblin et al., 1997) from which the participants in the Rice et al. (2004) single-gate study of kindergarten tense marking were drawn, 16.7% of the full study sample of 7,218 children resided in rural areas. This percentage is less than the 1990 census estimate as reported by Tomblin et al. (1997) of 25.4% of 5-year-old children living in rural areas as well as the 2010 census estimate of 19.3% of the total U.S. population living in rural areas. The county targeted for recruitment in the present study, therefore, represents a rural population that may have been underrepresented in the single-gate study of kindergarten tense marking. If this was indeed the case, then examination of kindergarten tense marking in a predominately rural cohort is needed to test whether the bimodality hypothesis holds up when tested in a previously underrepresented group.

The county targeted for recruitment can be considered economically disadvantaged. In 2014, the percentage of persons living in poverty in this county was greater than the national average (17.6% vs. 14.8%; U.S. Bureau of the Census, 2014a). The majority of students attending public schools in this county (64.6%) were considered economically disadvantaged due to their families meeting income requirements to receive free or reduced meals at school (Tennessee Department of Education, 2014). Educational attainment levels in this county lag behind national averages. Among persons 25 years of age, only 13.6% have a Bachelor's degree or higher. This figure contrasts with the TN state average of 24.4% and the national average of 29.3% (U.S. Bureau of the Census, 2010-2014). The economic and educational attainment status in this county, though concerning, was not considered a threat to the validity of this study. Despite the reported under-identification of language problems in children of lower socio-economic status (SES; Bishop & McDonald, 2009), the actual language profiles of low SES youngsters with language impairment are comparable to those of children with language problems from mid-high SES backgrounds (Roy, Chiat, & Dodd, 2014). Moreover, Rice et al. (1998) reported that maternal education level did not predict tense-marking growth over time among preschoolers with typical language or early school-aged children with SLI.

The school district targeted for recruitment also lies in a county that, according to data from the 2010 U.S. Census, is racially homogenous. The vast majority of county residents, 94.8%, identify themselves as White only (U.S. Bureau of the Census, 2014b). The school district is comprised of three elementary schools. The percentage of non-Hispanic White students at the three district elementary schools ranged from 92-97% (individually at 96.6%, 94.6% and 91.9%; Tennessee Department of Education, 2014). Because the effects of dialectical differences (e.g., African American English, Spanish-influenced English, Asian-

influenced English) and English Language Learner status on tense marking are not fully known, it was important to reduce bias by testing this skill in a population of predominantly Mainstream American English (MAE) speaking students. Moreover, the measure used to assess tense marking, the TEGI, was standardized on children who spoke MAE and came from homes where English was spoken at least 75% of the time (Rice & Wexler, 2001).

Because the school district is located in the rural south, it is not possible to eliminate the possibility that some participants were speakers of Southern White English (SWE) dialect. In fact, it was expected that this would be the case. The possible presence of SWE dialectical features in the language of targeted participants was determined to pose very minimal, if any, threat to the validity of the study design for two reasons.

First, the district lies in a county that, although rural, is geographically situated well west of the Appalachian Region (Appalachian Regional Commission, n.d.). As such, certain Appalachian English grammatical features, such as an overgeneralized singular form of past *BE* to plural subjects (e.g., *They was walking*), should not be prevalent in the district targeted for recruitment (Wolfram & Christian, 1976). Even if this feature were to be present in the language of some participants, it relates to auxiliary and copula BE subject-verb agreement and not the presence or absence of obligatory tense marking on lexical verbs.

Second, and more importantly, studies of SWE speakers have failed to demonstrate that frequent omissions of the PT and 3S tense morphemes assessed in the present study are a dialectical feature of SWE speakers with unimpaired language skills. In their examination of grammatical features in the spontaneous language samples of 19 six-year-old typical language speakers of a rural version of SWE, Oetting and McDonald (2001) reported infrequent omissions of PT and 3S markers. By contrast, the overall omissions of obligatory PT and 3S markers from

15 six-year-old SWE speakers with SLI from the same study were 1.7 to 6.7 times greater than their typical language SWE peers, with the greatest difference occurring for PT markers. Cleveland and Oetting (2013) further quantified some of the findings from Oetting and McDonald (2001) and reported a statistically significant difference in the mean percent obligatory 3S verbs marked for tense by typical SWE six-year-olds (93%) as compared to SLI SWE six-year-olds (71%; Cohen's d = 1.06). Accordingly, there is reason to suspect that the distributional pattern of kindergarten tense marking in SWE follows the same trend as that observed in MAE speakers. Therefore, the possible presence of SWE speakers in the present study was determined to pose a very minimal threat to validity.

Participants

All kindergarteners in each of the three elementary schools within one school district were invited to participate in a speech-language screening at the beginning of the 2014 - 2015 school year. If a child enrolled in the district after the date the screening invitation packets were sent home, the child was not invited to participate. Of the 203 screening invitation packets sent home in children's backpacks, 153 (or 75%) were returned with parent consent to participate. The rate of returned consent across the three elementary schools ranged from 73% - 83% per school. Five consented kindergartners were withdrawn from the study because they failed to meet eligibility criterion (see next paragraph). Thus, the participant sample included 148 kindergarten students, or 73% of the entire district kindergarten population. Of the final sample of 148 kindergarteners analyzed, 81 (54.7%) were boys. The mean age of the sample at the time of screening was 5;8 (SD = 5 months; Range = 4;11 – 6;10). All participants were assigned to general education kindergarten classrooms. Race/ethnicity was not collected on individual

participants; however, observation indicated the participant pool aligned with the county demographics (i.e., approximately 95% Caucasian).

Consented kindergartners were withdrawn from the study (n = 5) if one of several circumstances was met: (a) The child was not able to respond to the research tasks. One child was withdrawn because he was minimally verbal and was not yet functional with using an AAC device; he was the only consented child who was assigned to a resource classroom. (b) The child did not pass the TEGI Phonological Probe. This task assures that a child can consistently produce or approximate, in mono-morphemic words such as *bu<u>s</u>* and *be<u>d</u>*, the word final phonemes used to mark 3S (e.g., *Every day he paint<u>s</u>*) and PT (e.g., *Yesterday she clean<u>ed</u>*). Three children were withdrawn because they failed the TEGI Phonological Probe. (c) The child was not a native speaker of English. One child was withdrawn because the teacher confirmed the child was a native Spanish speaker with very limited English proficiency. (d) The child did not obtain a nonverbal IQ score within the typical range (i.e., standard score of 70 or above). No children were withdrawn based on this circumstance.

To ensure that students with potential linguistic vulnerabilities met basic criterion for nonverbal cognitive functioning, the 51 children who failed to meet the TEGI manual-recommended criterion scores[2] for either the 3S Probe, the PT Probe, or the TEGI Screening Test

[2] These criterion score cut points were developed by the Rice and Wexler (2001) to reflect, at each six-month age level between ages 3;0 – 8;11, at least 80% sensitivity in separating the distribution of the language-impaired group in the standardization sample from the normal group of their standardization sample. According to Rice and Wexler, "the rationale used to determine the cut points involved consideration of the bi-modal distribution of affectedness" (p. 36). For reasons related to the potential sampling bias in two-gate designs discussed above, it was not expected that the TEGI manual-recommended cut points would necessarily align with any cluster boundaries found in the current study. Specifically, children forming the language-impaired group in the TEGI standardization research sample were drawn from clinical caseloads. Clinicians were asked to refer children on their caseloads who were receiving language therapy. Documentation of language testing used to diagnose the language impairment was required, but the authors note that some of these language scores came from older testing conducted as many as 15 months prior. As Rice and Wexler (2001) point out, language progress in the

Score (average of 3S + PT) were administered the *Primary Test of Nonverbal Intelligence* (PTONI; Ehrler & McGhee, 2008). Of this subset, 42 children scored at least within the average range (standard score ≥ 85) and 9 scored in the low average range (standard scores between 70-84). Given evidence of comparably compromised academic, social participation, and subjective well-being outcomes between language-impaired children with at least average non-verbal IQ (NVIQ) and children with low-average NVIQ (Tomblin, 2008), all 51 students were included in the analyses.

Speech-language screening battery

In the TEGI 3S probe, children were shown 11 pictures (1 demonstration, 10 trials); each picture depicted a person engaging in an activity (e.g., teaching). The examiner provided a description of the picture (e.g., *This is a teacher*) and prompted the child to describe the action (*Tell me what she does*). The task is designed to elicit a simple sentence with a third-person singular subject to evaluate the child's production of the 3S tense marker in obligatory contexts (e.g., *She teaches*). Child responses were scored correct for inclusion of the 3S marker (with or without a subject produced; *she teaches* and *teaches* scored correct). Child responses were scored incorrect for omission of the obligatory 3S marker when a singular subject was used (e.g., *She teach*). In accordance with administration directions from the TEGI manual (Rice & Wexler, 2001), unmarked verbs in the absence of a sentential subject were re-prompted with, for

---

intervening time may have been sufficient such that, "if testing was completed today, the child may no longer qualify for the study" (p. 60). Additionally, in their description of the language-impaired standardization group, the TEGI authors disclose that children may have been included in this study "as a result of low performance on omnibus tests for reasons of low vocabulary or deficits in other areas of language that may not result in low performance on the grammatical markers tested on the [TEGI]" (p. 60). In sum, it may be that the TEGI manual-recommended criterion scores — because they were derived from the performance distribution of children in a language-impaired group, some of whom may have normalized and/or possessed intact tense marking — overestimate the boundary of a bi-modal distribution of kindergarten tense marking otherwise observed in a single-gate design, such as that in the current study.

example, *Remember, start with s/he*. An overall 3S Probe percent correct score was computed by dividing the number of scorable responses marked for 3S by the total number of scorable responses (max. = 10).

In the TEGI PT probe, children were shown 20 pairs of pictures (2 demonstration, 18 trials); the first picture in each pair of pictures depicts a person engaging in an activity. The examiner provided a description of the picture (e.g., *Here the boy is raking*). The second picture in each set depicts the activity completed. The examiner provided the information *Now he is done* and prompted the child to describe the completed action with *Tell me what he did*. The task is designed to elicit a simple sentence with a third-person subject to evaluate the child's production of the past tense in obligatory contexts (e.g., *He raked*). Similar to the scoring for 3S, child responses were scored correct for inclusion of the PT marker and incorrect for omission of the PT marker when a subject is used (e.g., *He rake*). Irregular PT verbs were scored correct for inclusion of a tensed form regardless of irregular marking (e.g., *She wrote*) or over-regularization (e.g., *She writed*). Following administration directions from the TEGI manual (Rice & Wexler, 2001), unmarked verbs in the absence of a sentential subject were re-prompted with, for example, *remember, start with s/he*. An overall PT Probe percent correct score was computed by dividing the number of scorable responses marked for PT by the total number of scorable responses (max. = 18). Scoring of both the TEGI 3S and TEGI PT probes were carried out following the guidelines delineated by Rice and Wexler (2001) in the TEGI manual.

Procedures

The participants were administered individually (in the following order) a screening battery consisting of the TEGI Phonological Probe, TEGI Third Person Singular Probe (3S) and TEGI Past Tense Probe (PT) and the Test of Articulation Performance – Screen (TAP-S: Bryant

& Bryant, 1983). Child responses to the items on the TEGI probes were orthographically transcribed on protocol forms at the time of administration; TAP-S responses were phonetically transcribed on protocol forms to indicate any error responses. Children who failed to meet the author-recommended criterion scores for the TEGI Screening Test, the individual 3S Probe, or the individual PT Probe were administered the PTONI at the end of the battery. Data collection was carried out by a team including the author (certified SLP) the PhD faculty director of the Vanderbilt Child Language and Literacy Lab (certified SLP), and a team of graduate research assistants (many of whom are certified SLPs). All team members read the TEGI manual and were trained in the standardized administration of the TEGI Phonological, 3S, and PT Probes by either the author or the faculty director prior to collecting data. The assessment team for every child tested included a lab member who was a certified SLP with experience working in elementary schools.

Derivation of variables

In accordance with TEGI manual scoring guidelines, for each participant the 3S and PT percent correct scores were averaged to generate a composite TEGI Screening Test Score (Rice & Wexler, 2001). Selection of the TEGI Screening Test Score as the primary variable of interest was done to promote the ecological validity of findings from the present study; the authors recommend clinical use of the TEGI Screening Test as a "valuable tool for large scale screening endeavors" to "quickly determine whether or not a child needs additional services" (Rice & Wexler, 2001, p. 8).

Creation of the composite TEGI Screening Test Score from the individual 3S and PT percent correct scores was psychometrically supported by reliability testing of the scores derived from the present study. A high Cronbach's coefficient alpha value of .907 was derived for the

3S and PT composite. This value exceeds the conservative .90 level recommended for scores on a scale where important decisions are made (Nunnally & Bernstein, 1994) and therefore, indicates high internal consistency reliability between scores on the individual 3S and PT probes. In other words, the high Cronbach's coefficient alpha offers strong evidence for the shared underlying construct, or domain, of tense proficiency hypothesized to be assessed by the two individual morpheme probes.

Another consideration in evaluating the reliability of scores obtained from a shared domain is reflected in the Spearman-Brown prophecy formula. This formula suggests that test reliability increases as a function of increased test items, provided that test items are drawn from a shared domain (Nunnally & Bernstein, 1994). The impact of random measurement errors is minimized in the context of increased test items. Accordingly, the individual percent correct scores of the TEGI 3S (10 items) and PT (18 items) probes were averaged into a composite to increase reliability and minimize the impact of random measurement error.

Scoring reliability

TEGI scoring was exhaustively checked. Every response on every protocol, as well as the calculation of percent correct scores, was double-scored by the author and trained graduate research assistants to ensure accurate coding of responses as correct or incorrect. The examiner who recorded the child's responses online did not double-score that child's responses. Scoring discrepancies were reviewed by a third examiner and then were resolved by mutual consensus between the double scorer and the third examiner. All PTONI scoring was checked in the same manner.

Reliability of online recording of child responses was carried out by a trained graduate research assistant who scored a random sample of 28% ($n = 41$) of the participants' TAP-S,

TEGI 3S Probe and TEGI PT Probe responses based on high fidelity audio recordings. Reliability scoring was performed on blank protocol forms and thus the procedure was blinded to the original online scoring. Agreement rates between the independent, blinded audio scoring and the aforementioned double-checked online scoring were 98% for 3S Probe scores, 96% for PT Probe scores, and 94% for the TAP-S Articulation Quotient scores.

CHAPTER III

RESULTS

To analyze whether the TEGI Screening Test Scores distributed non-normally, a Shapiro-Wilk Test of Normality was used. The histogram in Figure 4 illustrates the distribution. According to the Shapiro-Wilk Test, the distribution of TEGI Screening Test Scores deviated significantly from normality ($p < .001$).

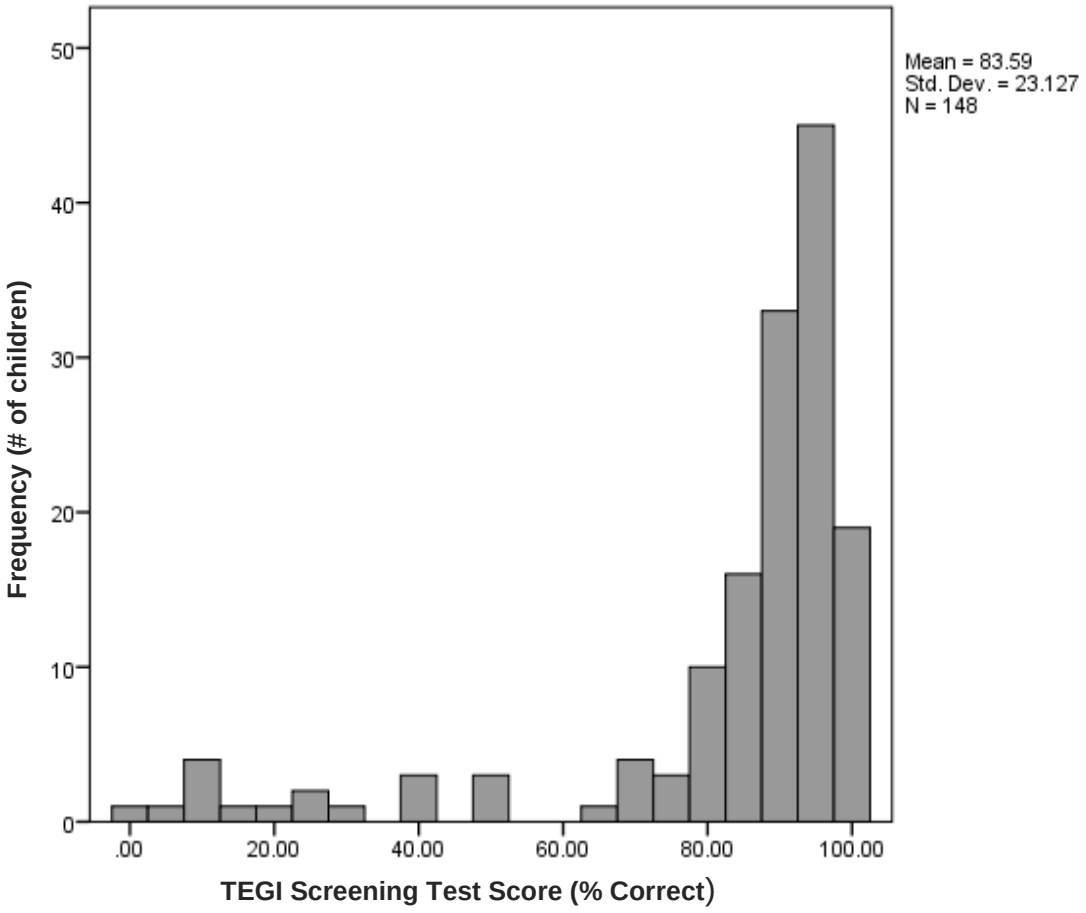

Figure 4. Distribution of TEGI Screening Test Scores (calculated as a mean of the Third Person Singular Probe Score and the Past Tense Probe Score; Rice & Wexler, 2001).

To explore the existence and number of latent classes potentially identifiable by the TEGI Screening Test Score data, a two-step cluster analysis was carried out in SPSS Statistics for Windows (Version 23). In a two-step cluster analysis, Ward's hierarchical method is applied initially to identify a logical cluster solution with good discriminatory power and minimal variance within each cluster. In the second step, a K-means iterative partitioning method makes multiple passes through the data, reassigning units from the first step to improve the accuracy of assignment to clusters (Hammet, van Kleeck, & Huberty, 2003). The automatically-generated best cluster solution is based on the Bayesian information criterion (BIC) for model selection among a finite set of models. BIC is a measure of goodness-of-fit with smaller values representing an increased fit (Mooi & Sarstedt, 2011). The model with both the smallest BIC is preferred (Norusis, 2010).

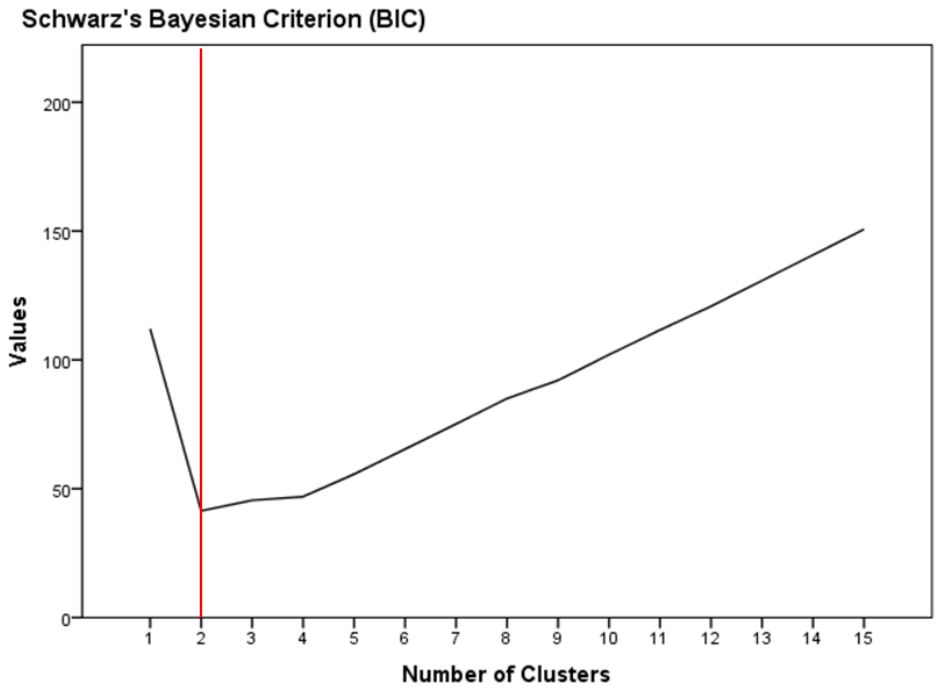

Figure 5. Cluster solutions and corresponding Bayesian information criterion (BIC) values. Vertical line illustrates the best-fitting cluster model resulting from the two-step cluster analysis (SPSS, Version 23).

As evidenced by the vertical line in Figure 2, a two-cluster solution is the best fitting model for these data. The cluster quality of this solution is supported by a high average silhouette coefficient for the entire data set. The silhouette coefficient is a helpful measure of the amount of clustering structure identified by the classification algorithm, in this case the two-step analysis. Silhouette coefficients reflect how well cases lie within their assigned cluster and are based on the dissimilarities of the Euclidian distances between cases within a cluster (Kaufman & Rousseeuw, 1990). Silhouette coefficients are dimensionless values that exist on a scale from -1 to 1, with values close to 1 representing "well classified" cases (e.g., the "within cluster" dissimilarity value is much smaller than the "between cluster" dissimilarity value) and values close to -1 representing "misclassified" cases. Kaufman and Rousseeuw (1990) proposed an interpretation for the average silhouette coefficient of an entire data set, illustrated in Table 1. The two-cluster model solution described above resulted in an average silhouette value of .84 (*SD* = .13), indicating strong cluster structure with good cohesion within and separation across clusters (see Figure 6).

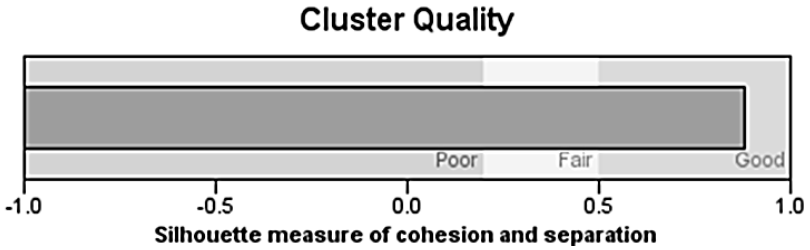

Figure 6. Cluster quality interpretation for the two-cluster model solution based on an average silhouette value of .84 (SPSS, Version 23).

Table 1

Interpretation of the Average Silhouette Coefficient for the Entire Data Set

Silhouette Coefficient	**Interpretation**
0.71 – 1.00	A strong cluster structure has been found
0.51 – 0.70	A reasonable cluster structure has been found
0.26 – 0.50	The cluster structure is weak and could be artificial; please try additional methods on this data set
$\leq$ 0.25	No substantial cluster structure has been found

Note. Based on Kaufman and Rousseew (1990). Copyright © 2005 John Wiley & Sons, Inc. Reprinted with permission.

As appreciated in Figure 7, the two-cluster solution offers evidence in support of tense marking as a bimodally-distributed skill in a general kindergarten population. The vast majority of cases (88.5%; n = 131) cluster around the upper end of proficiency (M = 90.91%, SD = 7.87; High Cluster). In contrast, the smaller cluster of cases (11.5%; n = 17) performed at or below approximately 50% accuracy on the TEGI Screening Test (M = 24.88%, SD = 17.93; Low Cluster; see Table 2 for descriptive statistics). Clear cluster separation is appreciated by the 12-percentage point gap between the lowest score in the High Cluster (63%) and the highest score in the Low Cluster (51%). The very large effect size difference between the two clusters (d = 4.77) provides validation of a non-arbitrary threshold delineating typical from atypical tense marking in kindergarteners.

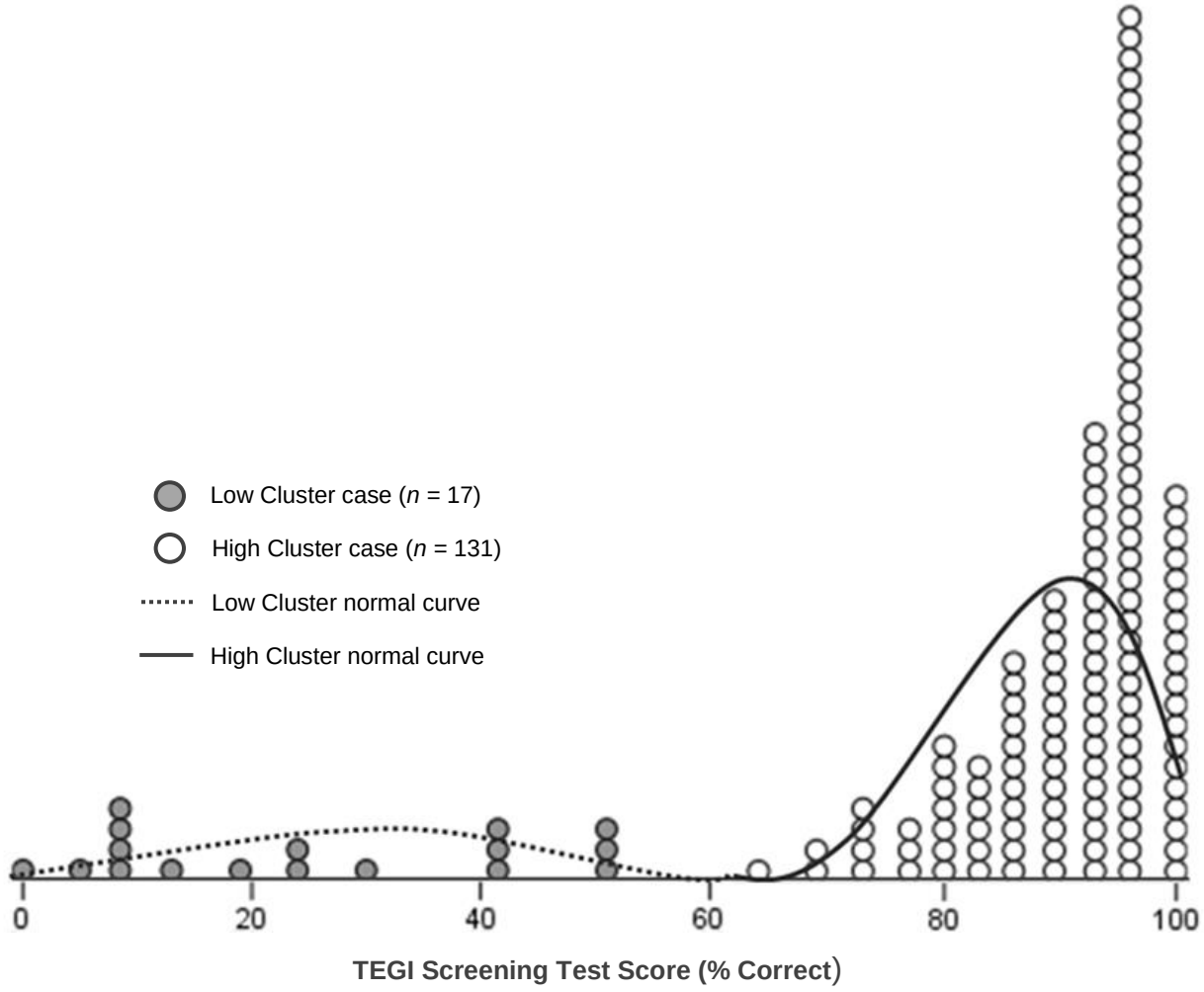

Figure 7. Distribution of individual TEGI Screening Test Scores (N = 148).

Differences in tense marking across the two clusters do not appear attributable to child chronological age or nonverbal intelligence as indexed by performance on the PTONI (Ehrler & McGhee, 2008; see Table 2). Cluster comparisons on these variables were non-significant. For the entire sample, TEGI Screening Test Scores (percent correct) were non-significantly correlated with chronological age (r = .12, p = .14) and PTONI standard scores (r = .001, p = .99). This pattern of non-significant correlations was found also for individual clusters (Low Cluster: TEGI Screening Test Score and chronological age (r = .32, p = .21), TEGI Screening

Test Score and PTONI standard score (r = .13, p = .63); High Cluster: TEGI Screening Test Score and chronological age (r = -.04, p = .67), TEGI Screening Test Score and PTONI standard score (r = -.07, p = .72). It is therefore highly unlikely that scores on the TEGI Screening Test were a proxy for chronological maturity or general cognitive level (Conti-Ramsden, Botting, & Farragher, 2001). A cluster difference in single word speech production accuracy, as measured by the TAP-S articulation screener, was noted (see Table 2). This finding is unsurprising given that the best estimate of speech delay prevalence (11%) in kindergarten-aged children with primary language impairment (i.e., nonverbal IQ $\geq$ 70) is almost three times greater than the overall prevalence of speech delay in six-year-old children (3.8%; Shriberg et al., 1999).

Table 2

Participant Characteristics and Testing Summary for Total Sample and by Cluster

Measure	Total Sample	Low Cluster[a]		High Cluster[b]	
	Mean (SD)	Mean (SD)	Range	Mean (SD)	Range
Age (months)	67.82 (5.12)	66.12 (4.48)	60 – 76	68.04 (5.23)	59 – 82
TEGI Screening Test Score (% correct)	83.59 (23.13)	25.12 (18.00)	0 – 51	91.18 (7.73)	64 – 100
PTONI SS	95.85[c] (14.69)	96.35[c] (16.20)	70 – 125	95.59[c] (14.12)	74 – 139
TAP-S AQ	95.69 (17.71)	78.06 (14.65)	< 58 –109	97.98 (16.80)	< 57 – 118

Note. SS = standard score; AQ = Articulation Quotient (similar to SS; *M* = 100; *SD* = 15); Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001); Primary Test of Nonverbal Intelligence (PTONI; Ehrler & McGhee, 2008); Test of Articulation Performance – Screen (TAP-S; Bryant & Bryant, 1983).

[a] *N* = 17 (7 girls, 10 boys). [b] *N* = 131 (60 girls, 71 boys).

[c] The PTONI was administered to all 17 children in the Low Cluster and 34 of the children from the lower tail of the High Cluster who scored below the TEGI manual-recommended criteria for the 3S Probe, the PT Probe, or the TEGI Screening Test (see Footnote 1).

Several tests of reliability were conducted to ensure that the two-cluster solution to the TEGI Screening Test data was accurate. First, a variation of split-half reliability was carried out by re-running the two-step cluster analyses with paired random halves of the TEGI Screener data (i.e., two randomized sets of 74 scores representing the full 148 score dataset). The results of both analyses were aligned highly with each other and with the original two-step analysis. In both of the half-samples, a two-cluster solution was found to best fit the data (see Figure 8). Moreover, the respective cluster sizes and individual cluster memberships were balanced across the two half samples and, when aggregated, mirrored exactly those from the full sample two-step cluster analysis findings (see Table 3).

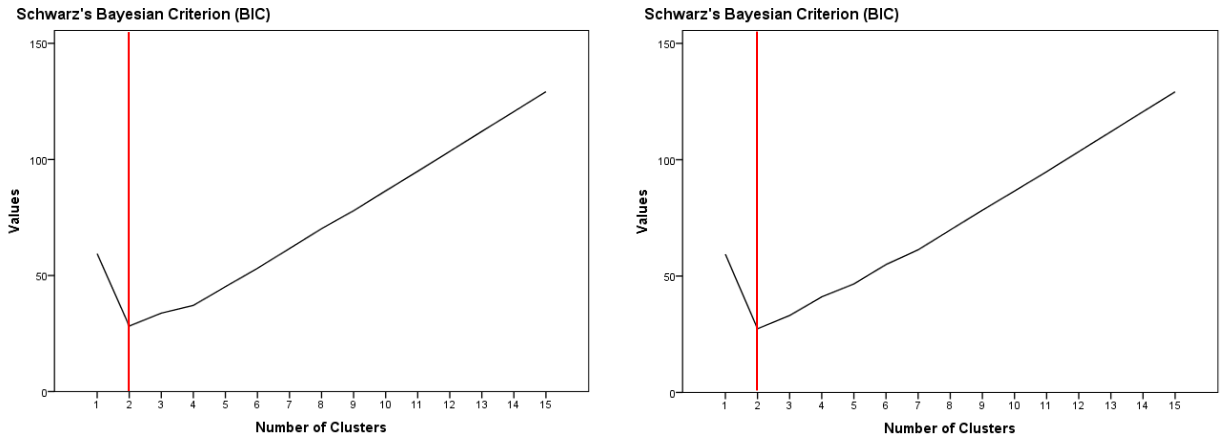

Figure 8. Cluster solutions and corresponding Bayesian information criterion (BIC) values for split-half reliability test. The BIC plot on the left is for the first half of the random sample (n = 74). The BIC plot on the second half of the random sample (n = 74; see Table 3). Vertical lines illustrate the best-fitting cluster model resulting from the two-step cluster analysis (SPSS, Version 23).

Table 3

Split-half Reliability Cluster Sizes and Members Relative to the Full Sample (N = 148)

	Low Cluster	**High Cluster**
First Half Random Sample (*n* = 74)		
Number of members	10	64
TEGI Screening Test[a] [M (SD)]	25.40 (17.46)	90.44 (8.57)
Second Half Random Sample (*n* = 74)		
Number of members	7	67
TEGI Screening Test[a] [M (SD)]	24.71 (20.16)	91.90 (6.81)
Full Sample (from Table 2)		
Number of members	17	131
TEGI Screening Test[a] [M (SD)]	25.12 (18.00)	91.18 (7.73)

Note. Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001). [a]Values represent percent correct scores.

Additional confirmation of the best fitting cluster solution for these data was carried out through visual inspection of the dendograms yielded from hierarchical agglomerative methods of cluster analysis using SPSS (Version 23). Agglomerative methods involve a series of successive mergers, or "linkages," of similar cases into groups (Aldenderfer & Blashfield, 1984). The analysis begins with each individual case representing its own cluster and ends with all cases subsumed under a single cluster. The sequence of successive mergers at each stage of the cluster analysis can be represented visually with a tree diagram, or dendogram. The "single linkage" hierarchical clustering method is one of the simplest agglomerative methods. The single linkage process searches for pairs of individual cases (or data points) based on "nearest neighbor"

distance (Everitt, Landau, & Leese, 2001). At each stage, a new candidate neighbor can be fused with an existing group on the basis of the highest level of similarity of any group member, hence the term "single linkage." The single linkage dendogram of the TEGI Screening Test Scores for all 148 cases (or children) in Figure 9 visually illustrates this hierarchical clustering technique. The vertical height represents that distance at which each fusion is made (Everitt et al., 2001).

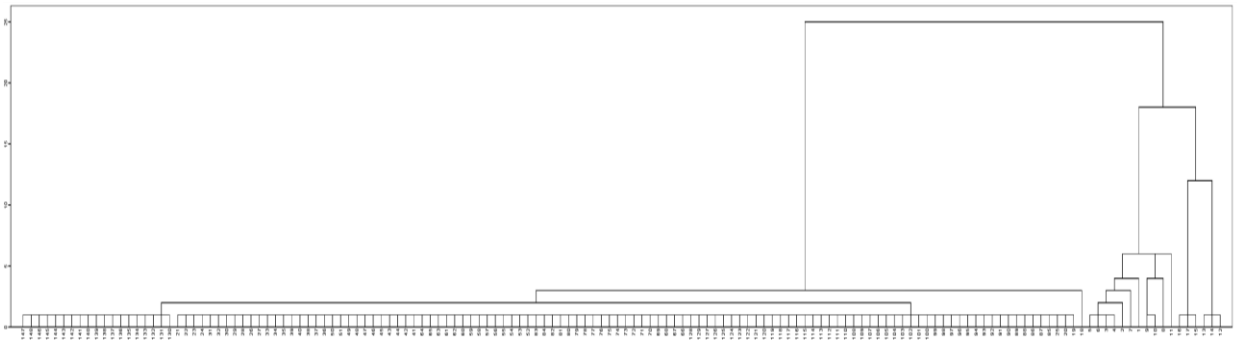

Figure 9. Single linkage dendogram for TEGI Screening Test Scores resulting from the single linkage hierarchical cluster analysis procedure (SPSS, Version 23). Each numerical tick mark on the *x*-axis represents an individual case (N = 148). The *y*-axis values represent the re-scaled distance units where clusters combine.

A drawback to the single linkage method is that cluster structure is not taken into account and thus, unbalanced chains of clusters are prone to emerge (Everitt et al., 2001). As such, determination of a hierarchical cluster solution through visual inspection of a dendogram is better carried out using Ward's method (Ward, 1963). Ward's hierarchical procedure maximizes between-cluster variability and minimizes within-cluster variability by calculating the squared Euclidean distance of each case to the cluster mean and then joining only those cases that result in small increases in the overall sum of squared within-cluster distances (Aldenderfer & Blashfield, 1984; Norusis, 2010). Determination of the number of clusters that best fit the data requires some interpretation. As a general rule, the minimum number of relatively cohesive clusters that account for as much of the data as possible is preferred (Schwartz & Conture, 1988).

This rule can be fulfilled by examining the dendogram for the cluster number associated with the largest vertical distance change in cluster fusion levels (Everitt et al., 2001). The dendogram for the TEGI Screening Test Score data illustrated in Figure 10 was created using Ward's hierarchical method and clearly shows that the two-cluster solution best satisfies this rule. The findings from Ward's hierarchical method support the two-cluster solution from the two-step cluster analysis described above. Morever, the cluster sizes (*n*s = 131, 17) and cluster case members are identical to the results of the two-step method.

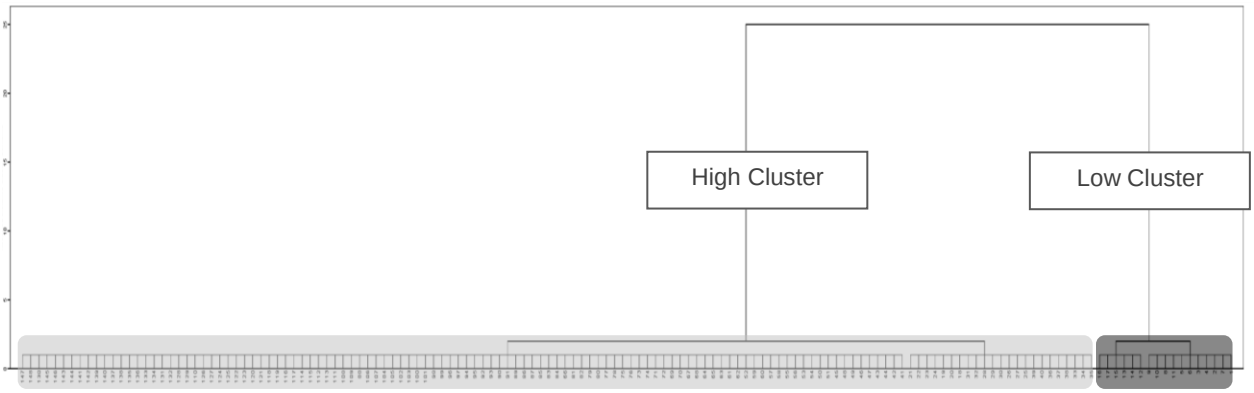

Figure 10. Ward linkage dendogram for TEGI Screening Test Scores created from Ward's method hierarchical cluster analysis procedure (SPSS, Version 23). Each numerical tick mark on the *x*-axis represents an individual case (*N* = 148). The *y*-axis values represent the re-scaled distance units where clusters combine. Light-shaded cluster = High Cluster (*n* = 131); Dark-shaded cluster = Low Cluster (*n* = 17).

A final reliability check for the two-cluster solution was carried out following two numerical criteria for interpreting Ward's hierarchical cluster analysis offered by Lambert, Brannan, Breda, Helfinger, and Bickman (1998). In selecting the ideal number of clusters to describe a sample, Lambert et al. noted that: "(a) A good clustering solution should have a higher R^2 than expected by chance clustering of random numbers [and] (b) The cubic clustering criterion (Sarle, 1983) should show a local peak indicating an optimal number of clusters" (p. 49, 1998). To apply these criteria to the present TEGI Screening Test Score data, the CLUSTER

Procedure for Ward's Minimum Variance Cluster Analysis was carried out using Statistical Analysis System software (SAS, Version 9.4). This analysis yielded R^2 values indicating the proportion of variance accounted for by the clusters. Additionally, an *approximate expected value of* R^2 under the null hypothesis — that the data have a uniform distribution instead of forming distinct clusters — is provided. Figure 11 plots, for each cluster solution, the difference between the actual R^2 value and the R^2 value expected by chance. As can be seen, the highest R^2 difference was found for the two-cluster solution.

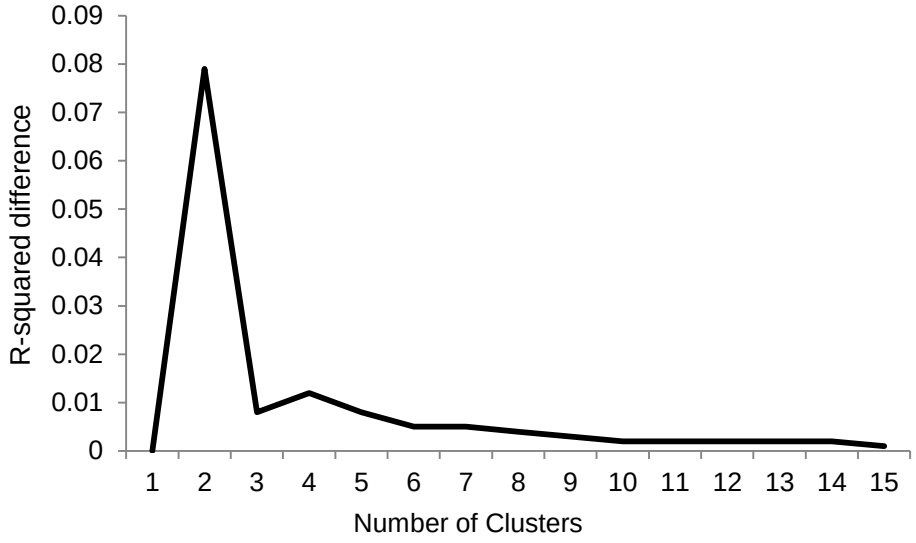

Figure 11. Difference between the actual R^2 value and the R^2 value expected by chance for each cluster solution (SAS, Version 9.4). The two-cluster solution, having the greatest R^2 difference, met Lambert et al.'s (1998) criteria for a good solution.

Figure 12 plots the cubic clustering criterion (CCC) statistic for estimating the number of clusters. Peaks in the plot of the cubic clustering criterion with values greater than 2 or 3 indicate good clusters; peaks with values between 0 and 2 indicate possible clusters (SAS, Version 9.4). There is a local peak of the CCC when the number of clusters is two. The CCC drops at three clusters and then steadily increases, surpassing the two cluster value again only at 15 clusters. For the sake of parsimony, solutions with fewer clusters are preferred (Lambert et

al., 1998). Therefore, in addition to the R^2 difference criterion, a two-cluster solution is supported additionally by the cubic clustering criterion.

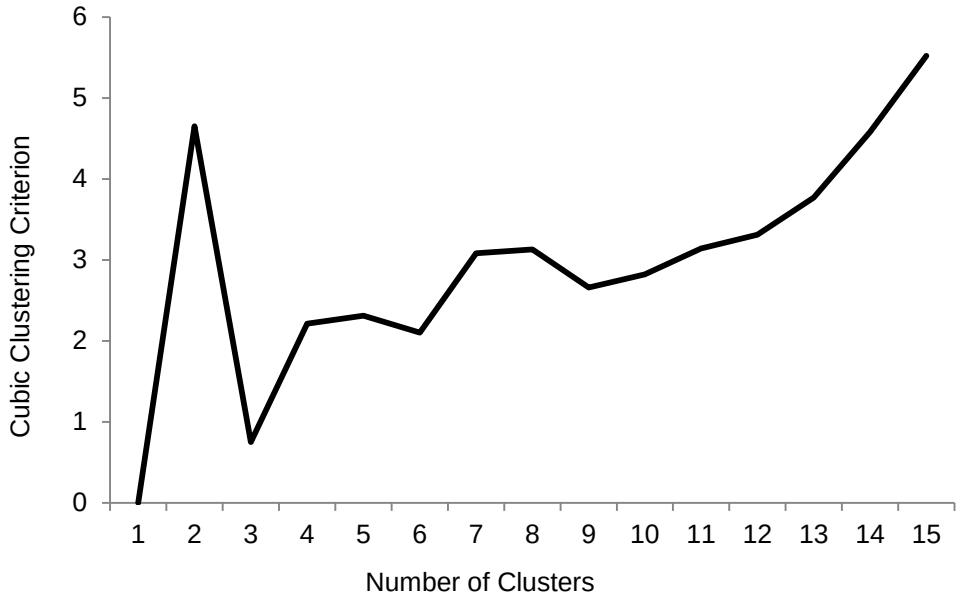

Figure 12. Cubic clustering criterion (CCC) statistic for estimating a cluster solution (Sarle, 1983; SAS, Version 9.4). The local peak CCC value at two clusters met Lambert et al.'s (1998) criteria for an optimal cluster solution.

CHAPTER IV

DISCUSSION

The aim of the present study was to test the generalizability, or transfer, of the bimodality hypothesis of kindergarten tense marking to a population-based cohort sample using a single-gate study design. A single-gate design was employed to minimize the threat of sampling or spectrum bias from two-gate designs perhaps operating in previous findings pointing toward a bimodal distribution of kindergarten tense marking proficiency. Only one other single-gate study of tense marking is known to exist. Rice and colleagues' (2004) study was conducted with data collected over 20 years ago. Additional epidemiological investigations of SLI — and the candidate clinical makers of this diagnosis — are long overdue (Redmond, 2016). Further investigations of the distribution of kindergarten tense marking are especially warranted considering that the results of Rice et al. (2004) leave unclear the generalization of a bimodal distribution to a large, unfiltered sample.

Findings from the present study therefore offer an important next step in elucidating the exact status of tense-marking proficiency in a general population of kindergarteners. Cluster analysis revealed a categorical structure underlying the distribution of this skill. The best fitting two-cluster solution appears well aligned with a previously posited bimodal distribution (Rice, 2000). Further validation of the distributional findings from the present study can be evaluated by considering whether the assumptions of transferability of test results are fulfilled. Irwig, Bossuyt, Glasziou, Gatsonis, and Lijmer (2002), in a discussion of designing medical studies to ensure that estimates of test accuracy are transferable — identified six such assumptions. The

first two assumptions — (1) *The definition of disease is constant* and (2) *The same test is used* — are readily satisfied in the present study. Strictly speaking, the "disease" under present consideration is reduced tense-marking proficiency. Practically speaking, reduced tense-marking proficiency in MAE-speaking kindergarteners signals an impaired aspect of language development which, in turn, raises suspicion of a clinical diagnosis of a "disease" state, in this case language impairment. Tense-marking proficiency was defined in accordance with the scoring guidelines from TEGI Screening Test manual (Rice & Wexler, 2001), which are themselves consistent with widely-established standards for operationally defining this skill (Leonard, 2014). The test used in the present study, the TEGI Screening Test, is the same as that used in Rice et al. (2004). Moreover, any test of productive tense marking accuracy should conceptually be "the same" (or at least highly similar) provided that it was designed to tap the construct of finiteness, which is true for the measures used in the majority of studies where tense marking was tested.

The remaining four assumptions (3 – 6) of transferability relate to the distribution of test results, which is most germane to the present study. Each of these four assumptions (see Figure 13) will be discussed relative to how well the results of the present study map onto our best distributional property estimates of tense marking in kindergarten children as well as the diagnosis of SLI.

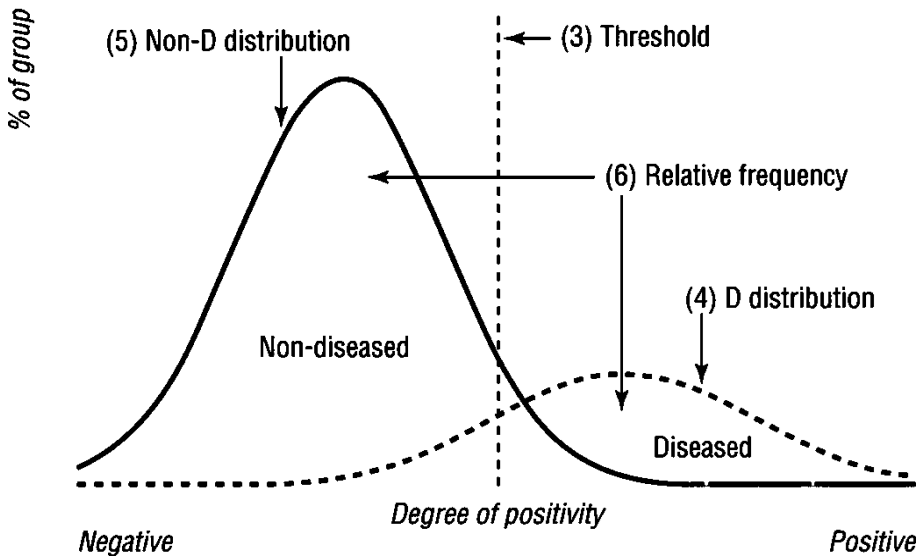

Figure 13. Distribution of test results in patients with and without the target disease. The numbers refer to assumptions for the transferability of test results (Irwig et al., 2002). Copyright © 2002 BMJ Publishing Group Ltd. Reprinted with permission.

(3) *The thresholds between categories of test results are constant.*

Central to the bimodal distribution hypothesis of tense marking among kindergarten children is the presence of a non-arbitrary threshold separating a category of children at the upper-end of the distribution from a category of children who cluster at the bottom of the distribution (Rice, 2000). As illustrated in Figure 1, in the study conducted by Rice and Wexler (1996), this threshold value appeared to reach maximum discriminant accuracy in separating 5-year-olds with SLI from age-matched peers with typical language at a score cut point of approximately 65-75% (Rice, 1998). Visual inspection of the two-cluster distribution in the present study suggests that a broadly similar threshold value of approximately 60% separates the two clusters (see Figure 7).

To more accurately calibrate a threshold cut point, a receiver operating characteristic (ROC) curve can be used (Irwig et al., 2002). ROC curves graphically represent the diagnostic accuracy tradeoffs between sensitivity and specificity at different thresholds. To generate an

ROC curve, an index measure (in this case, the TEGI Screening Test Score) must be compared to a reference standard. The design of the present study did not allow for administration of a reference standard measure to all the participants in the sample. As such, differential verification bias (i.e., bias introduced when the diagnostic status of only some of the participants in a sample is determined) was not controlled (Dollaghan, 2007).

Despite this limitation, a preliminary ROC curve can be generated using the diagnostic status of a subset of eight children from each of the two clusters who participated in additional language testing as part of the grant-funded study (Weiler, 2014). Participants for this grant-funded study were recruited based on failure to meet the author-recommended age-criterion TEGI Screening Test Score for the TEGI Screening Test (Rice & Wexler, 2001). The eight children from the High Cluster who participated are therefore represented in the bottom tail of that cluster's distribution (see Figure 7). As such, utilization of this subset of High Cluster children is considered a conservative approach to evaluating classification alignment with a reference standard. Diagnostic status was determined using a reference standard measure for language impairment with good discriminant accuracy. With a cutoff standard score of 95, Perona, Plante, and Vance (2005) used performance on the Structured Photographic Expressive Language Test: Third Edition (SPELT-3; Dawson, Stout & Eyer, 2003) to accurately identify children aged 4;0-5;10 as language impaired (LI; $n = 42$) or typically developing (TD; $n = 43$) with 90% sensitivity and 100% specificity. In their study, Perona et al. (2005) tested the diagnostic accuracy of the SPELT-3 — itself a test of morphosyntax — against a reference standard of clinical judgement by an SLP in combination with child performance on the Test for Examining Expressive Morphology (TEEM; Shipley, Stone, & Sue, 1983).

In the present study, child performance on the SPELT-3 is used as the reference standard to preliminarily determine the TEGI Screening Test Score threshold with the highest diagnostic accuracy. Regardless of their cluster membership (e.g., High vs. Low), participants were diagnosed LI if they scored below 95 on the SPELT-3 and TD if they scored at or above 95. Table 4 presents the individual participants' performance on the index measure (TEGI Screening Test Score) and the reference standard measure (SPELT-3) relative to their cluster assignment.

Table 4

Testing Summary by Cluster for Subset of Participants enrolled in Weiler (2014) Study

		Low Cluster							
Measure	**M (*SD*)**	**S1**	**S2[a]**	**S3[bc]**	**S4[ac]**	**S5[ac]**	**S6**	**S7**	**S8[b]**
TEGI Screening Test (% correct)	21.50* (*20.09*)	0	8	24	5	24	9	51	51
SPELT-3 SS	81.00* (*11.83*)	81	94	80	78	65	69	80	101
TAP-S AQ	73.00 (*16.56*)	74	79	67	57	67	73	109	57
PTONI SS	94.50 (*12.29*)	99	105	113	95	86	94	92	72
PPVT-4 SS	97.38 (*7.67*)	90	105	93	109	96	104	94	88
EVT-2 SS	95.88 (*11.24*)	98	97	93	121	91	96	86	85
		High Cluster							
Measure	**M (*SD*)**	**S9**	**S10**	**S11**	**S12**	**S13**	**S14**	**S15**	**S16**
TEGI Screening Test (% correct)	73.50* (*3.86*)	68	79	76	70	73	72	78	72
SPELT-3 SS	103.25* (*4.95*)	102	105	100	101	113	96	105	104
TAP-S AQ	85.20 (*12.85*)	99	87	90	88	66	87	99	66
PTONI SS	101.13 (*17.52*)	92	139	80	92	107	94	100	105
PPVT-4 SS	100.00 (*10.09*)	102	101	84	109	111	93	90	110
EVT-2 SS	100.88 (*11.24*)	86	110	89	95	118	96	103	110

Note. Between-cluster comparisons: *$p < .001$; All other between cluster mean comparisons *n.s.*; SS = standard score; AQ = Articulation Quotient; Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001); Structured Photographic Expressive Language Test: Third Edition (SPELT-3; Dawson, Stout & Eyer, 2003); Test of Articulation Performance – Screen (TAP-S; Bryant & Bryant, 1983); Primary Test of Nonverbal Intelligence (PTONI; Ehrler & McGhee, 2008); Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4; Dunn & Dunn, 2007); Expressive Vocabulary Test, Second Edition (EVT-2; Williams, 2007).

[a]Enrolled in speech and language therapy per parent report; [b]Enrolled in speech-only therapy per parent report; [c]Diagnosed with ADD or ADHD per parent report.

The sensitivity of the TEGI Screening Test two-cluster solution — that is, the proportion of children whose SPELT-3 scores met LI criteria who were assigned to the Low Cluster — is 100% (7/7). The specificity of the TEGI Screening Test two-cluster solution — that is, the proportion of children whose SPELT-3 score met TD criteria who were assigned to the High Cluster — is 88.9% (8/9). According to Plante and Vance (1994), discriminant accuracy above 80% is considered fair and accuracy above 90% is good. Following these guidelines, the TEGI Screening Test Score cluster membership in this subset of participants resulted in very good sensitivity for the diagnosis of LI and borderline-good specificity for the exclusion of LI.

Such diagnostic accuracy was corroborated through the calculation of likelihood ratios. "high" negative likelihood ratio of .01. Positive likelihood ratios represent confidence that an "affected" score on a test (in this case, membership in the Low Cluster) comes from a person who indeed has the target disorder (in this case, a language impairment as referenced by a SPELT-3 score under 95) as opposed to one without the disorder (Dollaghan, 2007). According to Sackett et al. (2000), positive likelihood ratios above 3 represent moderate confidence whereas those at or above 10 can be interpreted very confidently as indicating a disorder. A positive likelihood ratio of 9.09 was calculated, instilling confidence that Low Cluster classification indicates impaired language as referenced by SPELT-3 performance. On the flipside, negative likelihood ratios represent confidence that an "unaffected" score on a test (in this case, membership in the High Cluster) comes from a person free of the target disorder (in this case, typical language as referenced by a SPELT-3 above 95) as opposed to one with the disorder. Negative likelihood ratios at or below .10 suggest confidence that is it highly unlikely that an "unaffected" score came from someone with the disorder. A negative likelihood ratio of .01 was

calculated, instilling confidence that High Cluster classification indicates non-impaired language as referenced by SPELT-3 performance.

The actual threshold value that is reflected in the aforementioned psychometric properties for diagnostic classification can be evaluated using the ROC curve in Figure 14. The ROC curve for the TEGI Screening Test Scores is represented by the dashed, broken line. The connected, straight diagonal line represents classification accuracy values that are at chance (e.g., 50%). As a rule, the greater the discriminant accuracy of a test, the greater the area under the ROC curve. Perfect accuracy corresponds to an area of 1.0. For this ROC curve, the area under the curve is .992, which is significantly greater than chance ($p < .01$). Guidelines from Perkins and Schisterman (2006) for identifying the optimum threshold cutoff point on an ROC curve using the Youden Index are reported in Redmond, Thompson, and Goldstein (2011). The point at which the maximal vertical distance between the ROC curve and the diagonal reference line lies is considered the optimum threshold for discriminant accuracy. This point is represented by the star in Figure 14 and corresponds to a TEGI Screening Test Score of 59.5% (see Table 5). This value is generally constant with the cutoff value of approximately 65-75% appreciated in Figure 1 (Rice, 1998) and can therefore be cautiously interpreted at meeting the third assumption for the transferability of test results. The difference in threshold values between the present study and that reported in Rice (1998) may be attributed, in part, to differences in recruitment design (e.g., single-gate vs. two-gate), variations in the cohort characteristics, and/or the confidence interval around a given score. Although confidence intervals are not reported in the TEGI manual, Rice and Wexler (2001) did note that the mean absolute score differences under test-retest conditions were 7% and 8% for the TEGI 3S Probe and TEGI PT Probe, respectively.

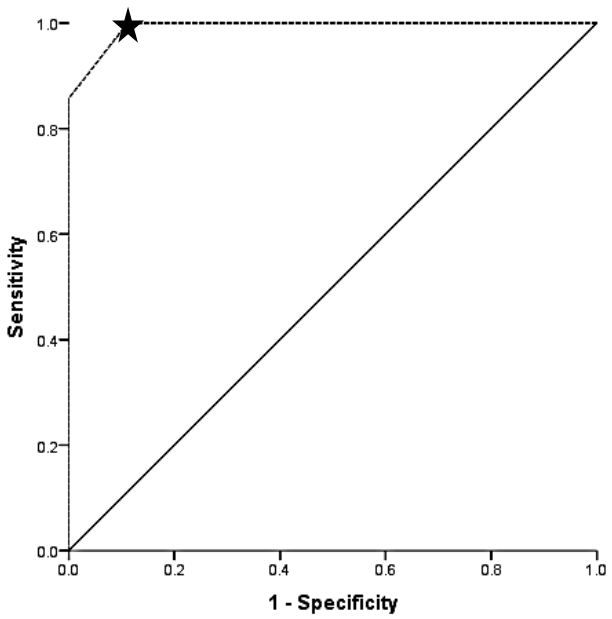

Figure 14. Receiver operating characteristic (ROC) curve for TEGI Screening Test Score classification accuracy (broken line). Straight diagonal line = chance classification accuracy. Star = optimum threshold score for classification accuracy (59.50%; see Table 5).

Table 5

Coordinates of the Receiver Operating Characteristic (ROC) curve

TEGI Screening Test Score	Sensitivity	1 - Specificity
2.50	.14	.00
6.50	.29	.00
8.40	.43	.00
16.05	.57	.00
23.65	.71	.00
37.50	.86	.00
59.50[a]	1.00	.11
68.50	1.00	.22
70.25	1.00	.33
71.75	1.00	.44
74.00	1.00	.67
76.75	1.00	.78
78.00	1.00	.89
79.50	1.00	1.00

Note. [a]Optimal cutoff point based on maximal sensitivity and specificity. Note the subtraction formula for specificity; actual specificity is .89.

(4) *The distribution of test results in the disease group is constant in average (location) and spread (shape).*

(5) *The distribution of test results in the group without the disease is constant in average (location) and spread (shape).*

A logical approach to testing assumptions (4) and (5) for the transferability of test results is to overlay the TEGI Screening Test Score distributions from the present study with tense composite score data from the study that inspired the bimodal distribution hypothesis (Rice & Wexler, 1996).

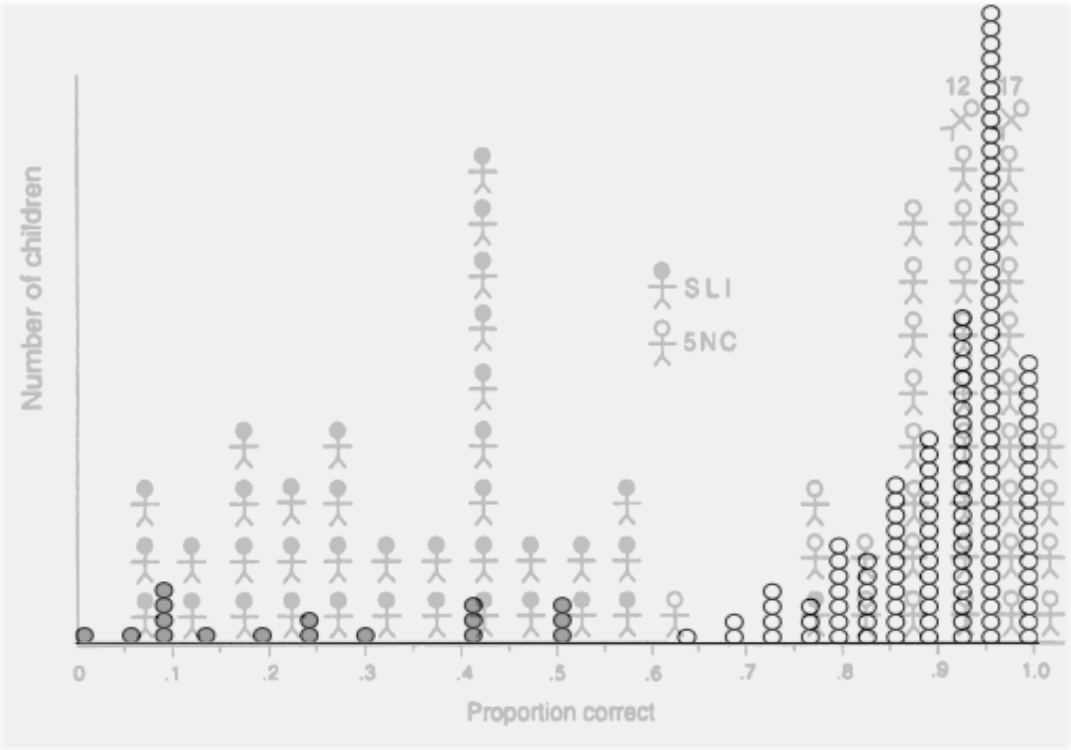

Figure 15. Overlay of TEGI Screening Test Score distribution from the present study (circles) with tense composite distribution from Rice and Wexler (1996; stick figures). See Figures 1 and 7 for details.

The main difference between the two sets of superimposed distributions illustrated in Figure 15 can be attributed to the nature of the study design. The matched group two-gate design of Rice and Wexler (1996) resulted in comparable numbers of children in the SLI and age-matched control (5NC) groups, respectively. In the present study, a single-gate approach predictably yielded a much smaller "affected" cluster (shaded circles) when compared to Rice and Wexler's SLI group but nonetheless in keeping with population prevalence estimates to be discussed shortly. Otherwise, the visually apparent constancy of "average" and "spread" across the two studies may be numerically evaluated by comparing means and standard deviations and the resultant effect sizes (Cohen's *d*). The very small to small-medium effect sizes for each of the groups or clusters fulfill the assumption of transfer of the distributional properties, for example, average (location) and spread (shape), across the two studies (see Table 6).

Table 6

Tense Composite Percentage Correct Score by Study Group/Cluster Comparison

	Group / Cluster	
Study	**SLI / Low Cluster M (*SD*)**	**5NC / High Cluster M (*SD*)**
Rice & Wexler (1996)[a]	32.6 (*16.20*)	90.7 (*5.80)*
Present	25.1 (*18.00*)	91.2 (*7.73*)
Effect size[b] *d*	*.44*	*.07*

Note. [a]Means and SDs based on Round 1 data reported in Rice et al. (1998). [b]Cohen (1988) interpretive benchmarks: small (*d* = 0.2), medium (*d* = 0.5), large (*d* = 0.8)

(6) *The ratio of disease to non-disease is constant.*

Not surprisingly, the gold standard to evaluating this assumption of transfer can be found in the prevalence estimates of kindergarten language impairment from Tomblin et al.'s (1997) epidemiological study. To carry out this evaluation, some preliminary — and tentative — conclusions must be drawn regarding the diagnostic status of the children comprising the Low-Performing Cluster as well as the High-Performing Cluster. Although, as described above, confirmatory testing was conducted on a subset of eight children from each cluster, there is no way to eliminate the potential for verification bias if generalizing those findings to the entire data set. Still, it seems reasonable to conclude that those children in the Low Cluster are phenotypically "impaired" in their development of an aspect of language that puts them at elevated risk for a language impairment diagnosis.

Children in the Low Cluster comprise 11.5% (17/148) of the study sample. To determine the constancy of language impairment prevalence estimates from the Tomblin study to the present findings requires a consideration of nonverbal IQ (NVIQ) functioning. Tomblin's 7.4% SLI prevalence finding was predicated on NVIQ inclusionary criteria of a standard score greater than 85. Practically speaking, the scoring of the NVIQ measure used in his epidemiology study — Block Design and Picture Completion subtests of the Wechsler Preschool and Primary Scale of Intelligence-Revised (Wechsler, 1989) — resulted in an actual "passing" standard score of greater than 87. Recall that the children in the Low Cluster (n = 17) from the present study obtained a range of nonverbal IQ scores (PTONI SS: 70 – 125). Twelve of these children scored above 87 on the PTONI. Applying Tomblin's NVIQ "passing" standard for an SLI diagnosis to the present study data set results in a prevalence rate of 8.1% (12 / 148). This 8.1% rate falls within the 95% confidence interval for the total SLI prevalence rate reported by Tomblin et al.

(6.3% - 8.5%; 1997) and offers evidence toward satisfying the final assumption of reliable and valid transferability of test results. The remaining five children from the Low Cluster, with NVIQs between 70 and 87, correspond to an overall prevalence rate of 3.4% (5/148). This rate generally aligns with the 5.1% prevalence rate for LI children with NVIQ scores in the same low-normal range from the Tomblin study, although no confidence intervals are reported (Rice et al., 2004). Finally, transferability of gender ratio was also observed. Tomblin et al. (1997) reported a 1.33:1 ratio of boys to girls in the SLI population. This ratio is constant with the 1.40:1 ratio of boys to girls in the Low Cluster with NVIQ > 87.

CHAPTER V

CLINICAL IMPLICATIONS AND FUTURE DIRECTIONS

Findings from this study provide another layer of evidence in support of the clinical marker utility of assessing tense marking in young school-aged children. The replication of a bimodal distribution of kindergarten tense marking in a single-gated population-based cohort should further confidence of the existence of a clear boundary separating children with typical development in this skill from children with impaired development in this skill. The latter group of children, as tentatively evidenced by confirmatory testing in the subset described in Table 4, should be considered as high-risk candidates for the diagnosis of language impairment. School-based SLPs and educators can readily harness this clinical marker by downloading the freely available TEGI Screening Test, along with the manual and stimulus pictures, at https://cldp.ku.edu/rice-wexler-tegi.

Current reports of the level of unidentified and untreated language impairment in elementary-aged children range from 45% (Redmond, Ash, & Hogan, 2015) to 54% (Bishop & McDonald, 2010). Although this rate of identification exceeds that from the Tomblin et al. (1997) epidemiological study (29%), we are still left with the bitter taste that as many as half of the elementary-aged children who meet research criterion for language impairment have yet to be identified. In a country that places a premium on the attainment of academic standards for all children, as represented perhaps most clearly in Congressional legislation like the No Child Left Behind Act of 2001, it is imperative that children with SLI be identified early in their educational

careers for services, especially because difficulties associated with SLI often persist into the adolescent years and beyond (see Nippold & Schwarz, 2002, for a review).

The increased likelihood of adverse effects on academic outcomes conferred by SLI is well-documented (cf. Tomblin & Nippold, 2014). In Poll et al. (2010), a single-gated study, the majority (62%) of adults with SLI reported a positive history of academic difficulties, especially in reading. Disturbingly, only one of the adults with SLI from Poll et al. reported ever having received language intervention. The authors interpret this finding as suggestive that the "absence of intervention cannot be attributed to the presence of mild impairments with no functional impact" (Poll et al., 2010, p. 425). Beyond academics, the impact of SLI may compromise the personal welfare and public safety of affected individuals. For example, elevated rates of self-reported physical bullying in children with SLI are three to four times higher than those reported by typical peers across the pre-adolescent years (e.g., Redmond, 2011; Conti-Ramsden & Botting, 2004) and over six times higher across the adolescent years (Knox & Conti-Ramsden, 2007). Additionally, young adults with SLI have been found to be limited in their understanding of Miranda rights (Rost & McGregor, 2012) as well as driving-related terminology from Driver's Manuals from the Department of Motor Vehicles (Pandolfe, 2015). Considering the pervading impact of SLI, high rates of under-identification are even more alarming.

At a small scale, the crisis of SLI under-identification was reflected in the present study. Among the subset of seven children described in Table 4 who participated in confirmatory testing and met the reference standard criteria for LI, only four (57%) were receiving speech and/or language services according to parent report. Consistent with prior reports of comorbid speech delay (e.g., Zhang & Tomblin, 2000; Bishop & McDonald, 2010) and/or ADHD diagnosis (e.g., Redmond et al., 2015) bolstering the likelihood of intervention receipt — perhaps

due to the more overtly recognizable behaviors associated with these conditions as compared to isolated LI — three of the four confirmatory group children receiving intervention in the present study had an ADD/ADHD diagnosis and all four had articulation difficulties sufficient to warrant speech therapy. The remaining three children meeting LI criteria (S1, S6, S7; Table 4) had language difficulties in the absence of ADD/ADHD or speech articulation difficulties sufficient to be flagged for therapy. Despite SPELT-3 scores well below the 95 cutoff for LI, none of these three children had ever been enrolled in language intervention. Were a diagnostician only to consider vocabulary skill, which, based on the results of Betz et al. (2013) discussed above is a distinct possibility, then none of these three children — nor any child from confirmatory subsample for that matter — likely would have presented as impaired as no PPVT-4 or EVT-2 standard score fell below 1 standard deviation (*SD* = 15) from the test mean score of 100.

The extent to which individual reports of unidentified language impairment in the elementary school population reflect national trends is currently indeterminable. Rates of SLI diagnoses and intervention enrollment are not tracked by either the Centers for Disease Control and Prevention, the U.S. Department of Education, the American Speech-Language-Hearing Association, or any other agency (Redmond, 2016). This knowledge gap should not discourage action. The results of the present study underscore the potential and feasibility for clinical use of the TEGI Screening Test in a kindergarten population for the purpose of identifying those children at risk for SLI who might very well fly under the radar otherwise. Given the brief 10-15 minute administration time for the TEGI Screening Test, it is not impractical to think that entire schools or school district might utilize a similar approach to that taken in the present study. If carried out in such a manner, local norms could be derived and used as the basis for screening pass/failure (e.g., Redmond et al., 2015). Findings from the present study may serve as a

preliminary reference point for the utilization of this approach. In particular, all 17 children in the Low Cluster fell at least 1.3 standard deviations below the total sample mean (M = 83.35, SD = 23.16) whereas all 131 children in the high cluster scored within .9 standard deviations — and all but one scoring within .7 standard deviations — of the total sample mean. These findings align with the local norms-based threshold of failing scores on the TEGI at or below the 10th percentile (i.e., -1.28 *SD*s) used by Redmond et al. (2015; as reported in Redmond, 2014)

Assessing tense marking as an alternative approach to identification — perhaps carried in combination with other known clinical markers of language impairment like sentence and nonword repetition (for a review, see Pawlowska, 2014) — has been advocated by other child language researchers, including Redmond (2016), who noted:

> An approach that has yet to be utilized in this area is basing initial SLI case assignment on children's performance on tense marking, nonword repetition, and sentence recall rather than basing them on potentially arbitrary standard score criteria from omnibus language tests. The advantage here would be better alignment in our prevalence and co-occurrence estimates with phenotypes of SLI currently being used in behavioral and molecular genetic investigations. (p. 20)

An added benefit to assessing tense marking for the purposes of identification is that it is more than just a flagging tool; its very definition (i.e., difficulties with consistent production of grammatical sentences) should also be a focus of intervention (Krok & Leonard, 2015). Indeed, tense-marking proficiency is clearly reflected in the Common Core State Standards Initiative currently adopted by 42 U.S. states. The English Language Arts Standards in Language for speaking and writing in grades kindergarten through third grade specifically note the use of verb tenses (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010).

Given the "illusory recovery" developmental trajectory of SLI (Scarborough & Dobrich, 1999; Stothard et al., 1998), we know that even though affected children use tense markers more

consistently in spoken language by the ages of eight or nine (Rice et al., 1998), they continue to regularly omit markers like past tense *–ed* in written language until at least age twelve (Windsor, Scott, & Street, 2000). In other words, although the outward symptoms associated with SLI may change over time, the underlying linguistic vulnerability persists, manifesting perhaps in different contexts. By the start of school entry in kindergarten, tense marking (i.e., finiteness) is an increasingly critical component of functionally communicative acts, for example presenting an oral report or writing about past events in a class journal, both of which are commonplace by first or second grade. It is therefore both psychometrically-sound and ecologically-valid to tap into tense marking for the purposes of improved identification of those children at risk for language-learning challenges.

What remains unclear — and what should be the basis for future studies in this area — is the extent to which a bimodal distribution of tense marking further transfers to the same cohort across school years or to other cohort populations that differ on such factors as geography, SES, dialect, age, and residential strata (e.g., urban, suburban, rural), to name a few. If a bimodal distribution of tense marking indeed holds up across variable populations and settings, then a further line of inquiry would be to characterize possible fluctuations in the threshold value (i.e., cutoff) between settings or populations and to examine the factor(s) underlying such potential differences. Oetting and colleagues have been trailblazers of such inquiries as they pertain to the diagnostic accuracy of measures used to identify SLI in AAE and SWE dialect speakers. For example, they found that the same empirically-derived sentence recall cutoff score was comparable in diagnostic accuracy across the two dialect populations (Oetting, McDonald, Seidel, & Hegarty, in press). Finally, in addition to continuing efforts to characterize the distributional properties of tense marking, child language researchers are encouraged to

empirically evaluate, through single-gated designs, the distributional structure of other clinical markers for pediatric SLI (e.g., nonword repetition, sentence recall). As seen in the present study, cluster analysis offers one possible approach for identifying subgroups based on performance on these and other measures.

CHAPTER VI

LIMITATIONS

Although a bimodal distribution of kindergarten tense marking was indicated in the present study, it cannot be assumed that this finding would transfer to other cohorts. Therefore, future studies in this area are recommended. Such studies would benefit from methodological considerations that address some of the limitations of the present study.

Even though parent permission return in the school district targeted for recruitment was high (75%), the possibility of different results with full (or closer to full) district participation cannot be ruled out. The sampling bias that may have resulted from the absence of a quarter of the kindergarten population in this district, however, seems acceptable considering the rate of consent return from large-scale, NIH-funded studies targeting grade-level, school-wide recruitment (e.g., 53.8%: Tomblin et al., 1997; 78%: Oetting, 2014).

The definition of SLI stipulates that the deficit in language ability cannot be attributed to hearing loss (Leonard, 2014). None of the 148 children tested had a visually apparent hearing aid or cochlear implant, nor did any display any behaviors during testing to suggest that they could not adequately hear the examiner. Additionally, none of the 16 children who participated in the confirmatory testing were reported by their parent to have a hearing loss. Still, it cannot be ruled out that some of the children in the present study may have had a hearing loss, particularly if it had been undetected at the time of data collection. Given prevalence estimates for mild or minimal unilateral or bilateral permanent hearing loss in the school-aged population ranging from 3.1% (Mehra, Eavey, & Kearny, 2009) to 5.2% (Bess, Dodd-Murphy, & Parker, 1998), it

is likely that a handful of children who participated in the present study may have been excluded had hearing status been assessed.

The presence of low nonverbal intelligence (e.g., below 70) is another exclusionary criterion for a primary language impairment (Leonard, 2014). In the present study, the nonverbal intelligence status was established for the approximately one-third of the sample with the lowest TEGI Screening Test scores. The nonsignificant cluster comparison on PTONI scores, therefore, should be viewed cautiously because it did not factor in for those children from the High Cluster who were not administered this measure. Future studies in this area should be designed such that, ideally, all participating children are administered a measure of NVIQ.

Similarly, to avoid verification bias when evaluating an index measure's diagnostic accuracy relative to a reference measure standard, all study participants should be administered both measures. As noted above, the present study was not designed to eliminate verification bias. Only eight children from each of the two clusters were administered the "gold standard" reference measure, the SPELT-3. Accordingly, and as described above, findings from the confirmatory testing for language impairment status in this study are preliminary and should be interpreted with caution.

Recall that despite all children (N = 148) having passed the TEGI Phonological Probe for marking of the final consonants /s, z, t, d/, the two clusters differed on the TAP-S Articulation Quotient. Each of the 31 items on the TAP-S is scored as either correct or incorrect based on production accuracy of the entire word. In other words, if any sound within a word is produced in error, the item is scored as incorrect. The presence of speech sound distortions with minimal, if any, impact on intelligibility, such as interdentalizations of /s/ and /z/, therefore result in the scoring of an item as incorrect. As a screening tool, the TAP-S captures speech production skill

at a broad level. Without a finer-grained consideration of the types of errors, however, it is difficult to make conclusions about the exact speech status (e.g., typical, delayed, developmentally appropriate errors) of the participants. Future studies in this area should thus include speech production measures that capture phoneme-level accuracy to better explain the possible linguistic interplay between grammatical tense marking and articulation skill.

Finally, even though there was not a convincing reason to expect an influence of dialect on the variable of interest (tense marking), this possibility cannot be entirely eliminated. Future studies may therefore benefit from the inclusion of language sampling so that dialectical features of individual participants can be systematically coded for (cf. Oetting & McDonald, 2001; Washington & Craig, 1994) and factored into analyses as indicated.

REFERENCES

Abel, A.D., Rice, M. L., & Bontempo, D. E. (2015). Effects of verb familiarity on finiteness marking in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 58,* 360-372.

Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis.* Beverly Hills, CA: Sage Publications.

Appalachian Regional Commission. (n.d.). *Counties in Appalachia.* Retrieved from www.arc.gov/counties

Ash, A., & Redmond, S. M. (2014). Using finiteness as a clinical marker to identify language impairment. *SIG 1 Perspectives on Language Learning and Education, 21,* 148-158.

Battaglia, M., Bucher, H., Egger, M., Grossenbacher, F., Minder, C., Pewsner, P. (2001). The Bayes library of diagnostic studies and reviews (2nd ed.). Kantonsspital Basel, Switzerland: University of Basel, Basel Insitute for Clinical Epidemiology. Available from daniel.pewsner@bluewin.ch.

Beitchman, J. H., Wilson, B., Johnson, C. J., Atkinson, L., Young, A., Adlaf, E., et al. (2001). Fourteen-year follow-up of speech/language-impaired and control children: Psychiatric outcome. *Journal of the American Academy of Child & Adolescent Psychiatry, 40,* 75-82.

Bess, F. H., Dodd-Murphy, J., & Parker, R. A. (1998). Children with minimal sensorineural hearing loss: prevalence, educational performance, and functional status. *Ear and Hearing, 19,* 339-354.

Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools, 44,* 133-146.

Bishop, D. V. M. (2004). Specific language impairment: Diagnostic dilemmas. In L. Verhoeven & H. van Balkmo (Eds.), *Classification of developmental language disorders: Theoretical issues and clinical implications* (pp. 309-326). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Bishop, D. V. M., & McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders, 44,* 600-615.

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

Bossuyt, P. M. & Leeflang, M. M. (2008). Chapter 6: Developing criteria for including studies. Deeks, J. J., Bossuyt, P. M., Gatsonis, C. *Cochrane handbook for systematic reviews of diagnostic test accuracy (Version 0.4; updated September 2008)*. The Cochrane Collaboration, 2008.

Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test*. Novato, CA: Academic Therapy Publications.

Bryant, B. R., & Bryant, D. L. (1983). *Test of Articulation Performance: Screen*: Austin, TX: Pro-Ed.

Cleveland, L. H., & Oetting, J. B. (2013). Children's marking of verbal–s by nonmainstream English dialect and clinical status. *American Journal of Speech-Language Pathology, 22*, 604-614.

Colorado Department of Education, (2010). *Colorado K-12 speech or language impairment guidelines for assessment and eligibility*. Denver, CO: Department of Education Exceptional Student Leadership Unit.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conti-Ramsden, G., & Botting, N. (2004). Social difficulties and victimization in children with SLI at 11 years of age. *Journal of Speech, Language, and Hearing Research, 47*, 145–161.

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

Conti-Ramsden, G., Durkin, K., Simkin, Z., & Knox, E. (2009). Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *International Journal of Language & Communication Disorders, 44*, 15-35.

Culatta, B., Page, J., & Ellis, J. (1983). Story retelling as a communicative performance screening tool. *Language, Speech, and Hearing Services in Schools, 14*, 66-74.

Dawson, J., & Stout, C. (2003). *Structured Photographic Expressive Language Test* (3rd ed.). DeKalb, IL: Janelle Publications.

Dollaghan, C. A. (2004). Taxometric analyses of specific language impairment in 3-and 4-year-old children. *Journal of Speech, Language, and Hearing Research, 47*, 464-475.

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

Dollaghan, C. A. (2011). Taxometric analyses of specific language impairment in 6-year-old children. *Journal of Speech, Language, and Hearing Research, 54*, 1361-1371.

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

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: Pearson Assessments.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test—Revised*. Circle Pines, MN: American Guidance Service.

Dunn, L.M., & Dunn, L. M. (1996). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Ehrler, D., & McGhee, R. (2008). *Primary Test of Nonverbal Intelligence*. Austin, TX: Pro-Ed.

Eickhoff, J.R., Betz, S. K., & Ristow, J. (2010, June). *Clinical procedures used by speech-language pathologists to diagnose SLI.* Paper presented at the Symposium on Research in Child Language Disorders, Madison, WI.

Everitt, B. S., Landau, S., & Leese, Mo. (2001). *Cluster analysis*. (4th. Ed.). London: Arnold.

Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 57*, 1363-1382.

Hammett, Lisa A., van Kleeck, A., & Huberty, C. J. (2003). Patterns of parents' extratextual interactions during book sharing with preschool children: A cluster analysis study. *Reading Research Quarterly*, *38*, 442-468.

Hammill, D. D., Brown, V. L., Larsen, S. C., &Wiederholt, J. L. (1994). *Test of Adolescent and Adult Language* (3rd ed.). Austin, TX: Pro-Ed.

Hoover, J. R., Storkel, H. L., & Rice, M. L. (2012). The interface between neighborhood density and optional infinitives: Normal development and specific language impairment. *Journal of Child Language, 39*, 835-862.

Irwig, L., Bossuyt, P., Glasziou, P., Gatsonis, C., & Lijmer, J. (2002). Designing studies to ensure that estimates of test accuracy are transferable. *British Medical Journal, 324*, 669-671.

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., et al. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42,* 744-760.

Kamhi, A. (1998). Trying to make sense of developmental language disorders. *Language, Speech, and Hearing Services in Schools, 29*, 35-44.

Knox, E., & Conti-Ramsden, G. (2007). Bullying in young people with a history of specific language impairment. *Educational and Child Psychology, 24,* 130–141.

Krok, W. C., & Leonard, L. B. (2015). Past tense production in children with and without specific language impairment across Germanic languages: A meta-analysis. *Journal of Speech, Language, and Hearing Research, 58,* 1326-1340.

Lambert, E. W., Brannan, A. M., Breda, C., Heflinger, C. A., & Bickman, L. (1998). Common patterns of service use in children's mental health. *Evaluation and Program Planning, 21*, 47-57.

Law, J., Reilly, S., & Snow, P. C. (2013). Child speech, language and communication need re-examined in a public health context: a new direction for the speech and language therapy profession. *International Journal of Language & Communication Disorders, 48,* 486-496.

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech, Language, and Hearing Research, 52,* 1401-1416.

Leadholm, B., & Miller, J**.** (1993). *Language sample analysis: The Wisconsin guide*. Milwaukee: Wisconsin Department of Public Instruction.

Leeflang, M. M. G, Bossuyt, P., & Irwig, L. (2009). Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *Journal of clinical epidemiology, 62*, 5-12.

Leonard, L. (1991). Specific language impairment as a clinical category. *Language, Speech, and Hearing Services in Schools, 22*, 66-68.

Leonard, L. B. (1994). Some problems facing accounts of morphological deficits in children with specific language impairments. In R. Watkins & M.L. Rice (Eds.), *Specific language impairments* (pp. 91-106). Baltimore, MD: Brookes.

Leonard, L. B. (2014a). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT press.

Leonard, Laurence B. (2014b). Specific language impairment across languages. *Child development perspectives, 8*(1), 1-5.

Mehra, S., Eavey, R. D., & Keamy, D. G. (2009). The epidemiology of hearing impairment in the United States: newborns, children, and adolescents. *Otolaryngology-Head and Neck Surgery, 140*, 461-472.

Miller, C. A., & Leonard, L. B. (1998). Deficits in finite verb morphology: Some assumptions in recent accounts of specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 701-707.

Mueller, K. L. (2012). Causation, correlation, or confound? *What the comorbidity of language impairment and ADHD can tell us about the etiology of these disorders.* Unpublished doctoral dissertation, University of Iowa.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). Common Core State Standards (English Language Arts Standards for Language). National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C.

Newcomer, P., & Hammill, D. (1988). *Test of Language Development—Primary* (2nd ed.). Austin, TX: Pro-Ed.

Nippold, M. A., & Schwarz, I. E. (2002). Do children recover from specific language impairment? *Advances in Speech-Language Pathology, 4*, 41-49.

Norusis, M. J. (2010). *PASW statistics 18 statistical procedures companion*. Upper Saddle River, NJ: Prentice Hall

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill.

Oetting, J. B. (2014, November). *Language impairment in children who speak nonmainstream dialects.* Paper presented at the American Speech-Language-Hearing Association's 24th Annual Research Symposium, Orlando, FL.

Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*, 207-223.

Oetting, J. B., McDonald, J. L., Seidel, C. M., & Hegarty, M. (in press). Sentence Recall by Children with SLI across Two Nonmainstream Dialects of English. *Journal of Speech, Language, and Hearing Research*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/26501934.

Pandolfe, J. M. (2015). A pilot study investigating comprehension of driving vocabulary in adolescents with language impairment. Unpublished master's thesis, University of Connecticut—Storrs.

Pawłowska, M.. (2014). Evaluation of three proposed markers for language impairment in English: A meta-analysis of diagnostic accuracy studies. *Journal of Speech, Language, and Hearing Research, 57*, 2261-2273.

Perkins, N. J., & Schisterman, E. F. (2006). The inconsistency of "optimal" cutpoints obtained using two criteria based on the receiver operating characteristic curve. *American journal of epidemiology, 163*, 670-675.

Perona, K., Plante, E., & Vance, R. (2005). Diagnostic accuracy of the structured photographic expressive language test: third edition (SPELT-3). *Language, Speech, and Hearing Services in Schools, 36*, 103-115.

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

Poll, G.H., Betz, S.K., & Miller, C.A. (2010). Indentification of clinical markers of specific language impairments in adults. *Journal of Speech, Language, and Hearing Research, 53*, 414-429.

Redmond, S. M. (2011). Peer victimization among students with specific language impairment, attention-deficit/hyperactivity disorder, and typical development. *Language, Speech, and Hearing Services in Schools, 42*, 520-535.

Redmond, S. M. (2014, June). *Markers, models, and measurement error: Exploring the links between attention deficits and language impairments.* Paper presented at the Symposium on Research in Child Language Disorders, Madison, WI.

Redmond, S. M. (2016). Markers, models, and measurement error: Exploring the links between attention deficits and language impairments. *Journal of Speech, Language, and Hearing Research*, 5*9*, 62-71.

Redmond, S. M., Ash, A. C., & Hogan, T. P. (2015). Consequences of co-occurring attention-deficit/hyperactivity disorder on children's language impairments. *Language, Speech, and Hearing Services in Schools, 46*(2), 68-80.

Redmond, S. M., Thompson, H., & Goldstein, S. (2011). Psycholinguistic profiling differentiates specific language impairment from typical development and from attention-deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry, 54*, 99-117.

Reilly, S., Tomblin, J. B., Law, J., McKean, C., Mensah, F. K., Morgan, A., et al. (2014). Specific language impairment: a convenient label for whom? *International Journal of Language & Communication Disorders, 49*, 416-451.

Reitsma, J. B., Rutjes, A. W. S., Whiting, P., Vlassov, V. V., Leeflang, M. M. G., Deeks, J. J.(2009). Chapter 9: Assessing methodological quality. Deeks, J. J., Bossuyt, P. M., Gatsonis, C. (eds.) *Cochrane handbook for systematic reviews of diagnostic test accuracy (Version 1.0.0)*. The Cochrane Collaboration, 2009. Available from http://srdta.cochrane.org/

Rice, M. L. (1998). In search of a grammatical marker of language impairment in children. *SIG 1 Perspectives on Language Learning and Education, 5*, 3-7.

Rice, M. (2000). Grammatical symptoms of specific language impairment. In D.V.M. Bishop & L.B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome*. Philadelphia, PA: Taylor & Francis Inc.

Rice, M. L. (2004). Growth models of developmental language disorders. In M. L. Rice & S. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 207-240). Mahwah, NJ: Lawrence Erlbaum.

Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., & Marquis, J. (2004). Grammatical tense deficits in children with specific language impairment (SLI) and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language, and Hearing Research, 47*, 816.

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

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1412-1431.

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

Rost, G. C., & McGregor, K. K. (2012). Miranda rights comprehension in young adults with specific language impairment. *American Journal of Speech-Language Pathology, 21*, 101-108.

Rousseeuw, P. J, & Kaufman, L. (1990). *Finding groups in data: An introduction to cluster analysis*. Hoboken, N.J.: John Wiley & Sons, Inc.

Roy, P., Chiat, S., & Dodd, B. (2014). *Language and socioeconomic disadvantage: From research to practice.* London: City University.

Ruscio, J., & Ruscio, A.M. (2004). A nontechnical introduction to the taxometric method. *Understanding Statistics, 3*, 151-194.

Sackett, D. L., Straus, S. E., Richardson, W. S., Glasziou, P., & Haynes, R. Br. (2000). *Evidence-based medicine: How to practice and teach EBM* (2nd ed.) . Edinburgh: Churchill Livingstone.

Sarle, W. S. *(1983). Cubic clustering criterion (SAS Technical Report A-108).* Cary, NC: SAS Institute Inc.

SAS (Version 9.4) [Computer software]. Cary, NC: SAS Institute Inc.

SAS Institute Inc. (2010*). SAS/IML® 9.22 User's Guide*. Cary, NC: SAS Institute Inc.

Scarborough, H., & Dobrich, W. (1990). Development of children with early language delay. *Journal of Speech and Hearing Research, 33*, 70-83.

Schwartz, H. D., & Conture, E. G. (1988). Subgrouping young stutterers: Preliminary behavioral bbservations. *Journal of Speech, Language, and Hearing Research, 31*, 62-71.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: The Psychological Corporation.

Shipley, K. G., Stone, T. A, & Sue, M. B. (1983). *Test for Examining Expressive Morphology.* Austin, TX: Pro-Ed.

.

Shriberg, L., Tomblin, J. B.., & McSweeny, J. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1461-1481.

SPSS Statistics for Windows (Version 23) [Computer software]. Armonk, NY: IBM Corp.

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

Spaulding, T. J, Szulga, M. S., & Figueroa, C. (2012). Using norm-referenced tests to determine severity of language impairment in children: Disconnect between US policy makers and test developers. *Language, Speech, and Hearing Services in Schools, 43*, 176-190.

Stothard, S., Snowling, M., Bishop, D. V. M., Chipchase, B., & Kaplan, C. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41*, 407-418.

Tager-Flusberg, H., & Cooper, J. (1999). Present and future possibilities for defining a phenotype for specific language impairment. *Journal of Speech, Language, and Hearing Research, 42*, 1275-1278.

Tennessee Department of Education, (2009). *Resource packet: Assessment of language impairment.* Nashville, TN: Author.

Tennessee Dartment of Education (2014a). *State report card.* Retrieved from https://www.tn.gov/education/topic/report-card

Tomblin, J. B. (2008). Validating diagnostic standards for specific language impairment using adolescent outcomes. In C.F. Norbury, J.B. Tomblin, & D.V.M. Bishop (Eds.), *Understanding developmental language disorders: From theory to practice*. New York: Psychology Press.

Tomblin, J. B., & Nippold, M. A. (2014). *Understanding individual differences in language development across the school years*. New York: Psychology Press.

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

Tomblin, J. B., 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.

Tomblin, J. B., & Zhang, X. (1999). Language patterns and etiology in children with specific language impairment. In H. Tager-Flusberg (Ed.), *Neurodevelopmental disorders* (pp. 361-382). Cambridge, MA: MIT Press.

U.S. Bureau of the Census (2010). *Percent urban and rural in 2010 by state* [Data file]. Retrieved from https://www.census.gov/geo/reference/ua/urban-rural-2010.html

U.S. Bureau of the Census (2010-2014). *American community survey, 5-year estimates.* Retrieved from http://www.census.gov/quickfacts/table/IPE120214/00,47085,47

U.S. Bureau of the Census (2014a). *Current population survey, annual social and economic supplement; Small area income and poverty estimates*. Retrieved from http://www.census.gov/quickfacts/table/IPE120214/00,47085

U.S. Bureau of the Census (2014b). *Population estimates program*. Retrieved from http://www.census.gov/quickfacts/table/IPE120214/00,47085,47#headnote-js-a

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*, 236–244.

Washington, J., & Craig, H. (1994). Dialectal forms during discourse of poor, urban, African American preschoolers. *Journal of Speech and Hearing Research, 37*, 816-823.

Watkins, R. V, & Johnson, B. W. (2004). Language abilities in children who stutter: Toward improved research and clinical applications. *Language, Speech, and Hearing Services in Schools, 35*, 82-89.

Wechsler, David. (1989). *Wechsler Preschool and Primary Scale of Intelligence—Revised*. New York: Psychological Corporation.

Weiler, B. (2014). [Participle-ed: The role of argument structure and interpretation]. Unpublished raw data.

Williams, K.T. (2007). *Expressive Vocabulary Test* (2nd ed.). Bloomington, MN: Pearson Assessments.

Windsor, J., Scott, C., & Street, C. (2000). Verb and noun morphology in the spoken and written language of children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*, 1322-1336.

Wolfram, W., & Christian, D. (1976). *Appalachian speech*. Arlington, VA: Center for Applied Linguistics.

Zhang, X., & Tomblin, J.B. (2000). The association of intervention receipt with speech-language profiles and social-demographic variables. *American Journal of Speech-Language Pathology, 9*, 345-357.

Zimmerman, I., Steiner, V., & Pond, R. (2002). *Preschool Language Scale* (4th ed.). San Antonio, TX: Psychological Corporation.