JSLHR

Research Article

Tense Marking in the Kindergarten Population: Testing the Bimodal Distribution Hypothesis

Brian Weiler[a] and C. Melanie Schuele[b]

Purpose: The purpose of this study was to explore whether evidence for a bimodal distribution of tense marking, previously documented in clinically referred samples, exists in a population-based sample of kindergarten children from a rural county in Tennessee.
Method: A measure of tense marking, the Test of Early Grammatical Impairment (TEGI) Screening Test, was individually administered to consented kindergarten students (N = 153) across three elementary schools in a single school district. The consented children constituted 73% of kindergartners in the district. Cluster analysis was used to evaluate the number and composition of latent classes that best fit the distribution of the TEGI Screening Test scores.
Results: Analysis of the scores revealed a distribution that deviated significantly from normality. Cluster analyses (Ward's, *k*-means, single linkage) revealed a two-cluster solution as the best fitting model. The very large effect-size difference in mean TEGI Screening Test score between the two clusters (d = 4.77) provides validation of an identifiable boundary delineating typical from atypical tense marking in this sample of kindergartners. The difference in tense marking across the two clusters was not attributable to child chronological age. The percentage of the sample comprising the low-performing cluster aligns with specific language impairment and developmental language disorder prevalence estimates.
Conclusion: Additional demonstrations of a bimodal distribution of tense marking in future studies with carefully defined samples could strengthen the clinical marker evidence and utility of this linguistic feature.

Children with specific language impairment (SLI), the most common type of developmental language disorder (DLD), comprise 7%–8% of the kindergarten population (Norbury et al., 2016[1]; Rice, 2020; Tomblin et al., 1997). These children, who demonstrate language difficulties not attributable to low nonverbal intelligence, hearing loss, or neurological damage (Leonard, 2014), are notoriously under-identified and, by extension, underserved. In Tomblin et al.'s (1997) epidemiological study, parents of only 29% of the 216 children meeting SLI research criteria had been previously informed that their child had a speech or language problem. Among those kindergartners 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 comorbidity of language impairment and speech delay in a subsample of this same study population of 6-year-olds (n = 1,328) was only 1.3% (Shriberg et al., 1999), we are left with the sobering message that kindergartners with SLI, the majority of whom have unremarkable speech articulation, are frequently unidentified. In more recent population-based studies, the rate of clinical identification of SLI in elementary-age children has not improved appreciably since the 1997 study by Tomblin et al. (45%, 32%, 25%, and 54%, respectively, Bishop & McDonald, 2009; Norbury et al., 2016; Oetting et al., 2016; and Redmond et al., 2015).

Under-identification might not be troublesome if SLI in kindergarten children was a fleeting phenomenon that later resolves. Quite to the contrary, these children, followed longitudinally, continue to lag behind peers with typical language (TL) through 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

[a]Department of Communication Sciences and Disorders, Western Kentucky University, Bowling Green
[b]Vanderbilt University Medical Center, Nashville, TN
Correspondence to Brian Weiler: brian.weiler@wku.edu

Editor-in-Chief: Stephen M. Camarata
Editor: Mary Alt
Received June 10, 2020
Revision received August 21, 2020
Accepted November 13, 2020
https://doi.org/10.1044/2020_JSLHR-20-00335

[1]Using Tomblin et al. (1997) criteria for SLI.

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

academic, behavioral, psychosocial, and vocational outcomes have been documented with multiple samples of children with SLI (Conti-Ramsden & Durkin, 2012; Conti-Ramsden et al., 2009, 2019; Law et al., 2009; Stothard et al., 1998) and general language impairment (Beitchman et al., 2001; Brownlie et al., 2016; Johnson et al., 2010, 1999) followed longitudinally.

Clearly, if the academic, behavioral, and socioemotional disadvantages conferred to children with SLI are to be minimized through the receipt of services (via special education and related services), these children have to be identified first. 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) that called for continued research to identify clinical markers of SLI (Tager-Flusberg & Cooper, 1999). Generally speaking, a clinical marker is a particular symptom or sign that carries diagnostic accuracy in determining the presence or absence of a disorder or disease (Poll et al., 2010). A clinical marker of SLI, then, could be considered performance on a task that is characteristic of and especially sensitive to the diagnosis of disordered language. The NIH panel noted that, for speakers of English, a "composite reflecting children's degree of use of several finite verb-related morphemes in obligatory contexts seems to hold considerable promise…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 *agreement* features that are obligatory in the matrix clause (Rice, 2004). The morphemes that mark finiteness can be freestanding, as is the case with the *BE* copula and auxiliary (e.g., "Emma is happy," "Quinn and Lillian are playing") and irregular past tense (PT; e.g., "Courtney ran"). Other finiteness-marking morphemes, such as regular PT *–ed* (e.g., "Quinn jumped") and third-person singular (3S) present tense *–s* and *–es* (e.g., "Emma laughs," "Lillian kisses"), are affixed to lexical verbs. Some finiteness markers, such as PT, carry only the tense feature of the clause. Others, such as 3S present tense, 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, in the preschool and early school–age years, of grammatical well-formedness (Rice, 2004). For the purposes of brevity and consistency with common clinical and research nomenclature, the term *finiteness marking* or *finiteness markers* will be referred heretofore as *tense marking* or *tense markers*.

Rates of obligatory tense marker omissions (e.g., **She __ running*, **Yesterday he play__*; the asterisk (*) denotes agrammatical) reliably distinguish children with SLI aged 3–8 years from same-age peers with TL. Over a dozen studies with mainstream American English (MAE) speakers have reported a 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 MAE-speaking SLI and TL groups has been reported longitudinally (Rice et al., 1998) and across data collection methods, including conversational samples as well as sentence elicitation tasks (e.g., Krok & Leonard, 2015; Rice & Wexler, 1996) and sentence recall tasks (e.g., Abel et al., 2015; Hoover et al., 2012).

Diagnostic Dilemmas and the Bimodal Distribution Hypothesis of Kindergarten Tense Marking

Frequently, clinical decision making concerning diagnosis and eligibility operates under the assumption of normality under a unimodal distribution. Common diagnostic and service eligibility standards often dictate, for example, that scores 1.5 *SD*s below a normative mean signify the presence of a delay or disorder (e.g., Colorado Department of Education, 2010; Tennessee Department of Education, 2018). Use of eligibility cut-points, whether 1, 1.5, or 2 *SD*s below a mean, runs the risk of arbitrarily dichotomizing a continuous metric when applied to a normally distributed skill (Bishop, 2014; Tomblin et al., 1997). Is there really a meaningful difference in the likelihood of the presence of functional language impairment between a child who scores 1.4 *SD*s below the mean and a child who scores 1.6 *SD*s below the mean? Instead, if kindergarten tense marking is indeed a bimodally distributed skill, it should allow for easier and more valid identification because a clear(er) boundary would separate the performance of children with language impairment from that of same-age peers with TL.

Instead of being distributed normally (i.e., a unimodal "bell curve" distribution), tense-marking proficiency at the point of school entry (i.e., kindergarten) has been hypothesized to follow a bimodal distribution; children with SLI cluster at the lower end of the distribution whereas children with TL 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 difficulty of identifying the boundary with which one determines "affectedness" (Spaulding et al., 2006, 2012).

Taxometric methods can refine an understanding of how tense-marking proficiency may be bimodally represented in a kindergarten population. 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). A bimodal distribution would be considered categorical, and a unimodal distribution would be considered continuous. Cluster analysis was employed in this study as a taxometric method to evaluate the extent to which kindergarten children indeed "cluster" within a bimodal distribution when the construct of finiteness is assessed via tense-marking accuracy.

Compelling evidence for the bimodal model comes from a study conducted by Rice and Wexler (1996) and further reported in the work of 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. Children in the SLI group averaged 58 months in age (range: 52–68 months), and those in the TL group averaged 60 months in age (range: 52–67 months). Rice and Wexler (1996) reported that 36 of the 37 children in their SLI group marked tense in obligatory contexts with less than 60% accuracy, whereas all but one of the 45 children in the normal language control group marked tense with approximately 80% or greater accuracy (see Figure 1).

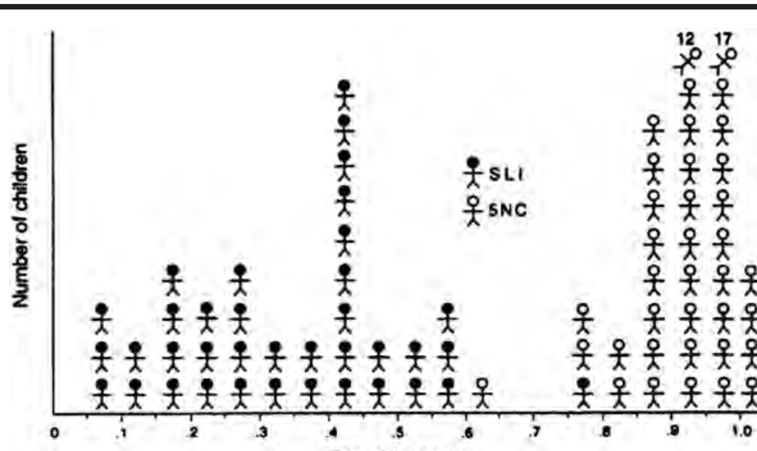
Cluster separation is appreciated by the practically nonoverlapping "buffer zone" of at least 10 percentage points (i.e., 65%–75%) between the two group distributions illustrated in Figure 1. The degree of nonoverlap between the two distributions, calculated by converting the Cohen's *d* effect size into 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 same-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 in children should be minimal.

The current project endeavored to address a limitation in the research methods used in studies indicating a bimodal distribution of tense. 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 verb tense morphology come from matched-group designs composed of an SLI group compared to a typically developing control group. They noted that studies "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%" (Reilly et al., 2014, p. 426).

Given the aforementioned report of only 29% clinical identification of SLI in the kindergarten population (Tomblin et al., 1997), it is quite possibly the case that the language impairment samples in matched-group designs using clinically ascertained samples were primarily composed of children with more severe deficits rather than children whose language deficits encompass the range. Arguably, children with more severe deficits are more likely to be 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 mild-to-moderate SLI, on the other hand, may very well be underrepresented 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 might 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 0.5–2 *SD*s above the population mean on measures of language skills. The potential for control group sampling bias is also reflected in a meta-analysis that reported an average 0.7-*SD* nonverbal IQ (NVIQ) advantage for age-matched peers when compared to children with SLI

Figure 1. Distribution of individual children's performance on a composite tense-marking score: specific language impairment (SLI) and age controls. 5NC = 5-year-old normal language controls. Reprinted with permission from Rice (1998).



(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-group design studies pointing toward a bimodal distribution of tense-marking proficiency, therefore, can be called into question. 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. Indeed, the authors of a study of tense-marking proficiency in a large population-based community sample of 6-year-olds (N = 676) reported a unimodal distribution of finite verb morphology composite (FVMC) percent accuracy scores derived from child conversational language samples with adult family members (Rudolph et al., 2019). They interpreted this distribution of scores, which was left-skewed (M = 93%, SD = 10), as lacking in clear evidence of different peaks for subgroups of children such as those with SLI. The unimodal FVMC distribution reported by Rudolph et al. (2019) has been offered as potential evidence for a dimensional (continuous) characterization of the linguistic deficits in SLI as compared to a taxonic (categorical) characterization (McGregor et al., 2020).

Several factors from the Rudolph et al. (2019) study, however, warrant consideration before closing the door on the need for further investigation on the distributional structure of tense marking in kindergarten-age children. Notably, there was not full consideration of the potential dialects spoken by the children in their sample and the impact of dialect on FVMC scoring. In nonmainstream dialects such as African American English (AAE), there is a higher percentage of grammatical zero forms for PT (e.g., *Yesterday she walkØ*) and 3S (e.g., *Everyday she eatØ*) when compared to MAE (e.g., Seymour et al., 1998; Washington & Craig, 1994). Of the 676 children whose FVMC scores were included in Rudolph et al.'s histogram plot, 30% were demographically classified as African American. Although the authors did adjust conventional FVMC scoring to accommodate two non-MAE dialect forms related to *BE*, there were no reported dialect modifications to other FVMC structures such as PT and 3S. The FVMC accuracy scores were calculated by dividing the number of morphemes overtly marked by the number of MAE obligatory contexts. It is therefore possible that lower percent accuracy scores in the distribution reflect MAE-speaking children with ungrammatical tense omissions and AAE-speaking children with grammatical zero forms. Partial evidence of the effect of nonmainstream home dialect on FVMC scores is reflected in Rudolph et al.'s report that the overt tense-marking percentage of the adult family member who conversed with the child was the strongest predictor, in turn, of the child's percent overt marking (R^2 = 18%). Additionally, there were no reported measures in the Rudolph et al. study to control for the possible omission of word-final tense morphology due to a child's phonological limitations. A phonological probe ensuring that a child can mark word-final phonemes carrying surface tense marking (e.g., /t, d/ for regular PT; /s, z/ for 3S) in monomorphemic words (e.g., "hat," "road," "mouse," "nose") would address this limitation. Finally, although descriptive statistics for the distribution of FVMC were provided, taxometric analyses such as cluster analysis offering a more refined examination of potential subgroups in the distribution were not reported. Rudolph et al. recommended the need for further investigation of proposed clinical markers, such as tense-marking proficiency, that are "representative of the conditions in which such markers will be employed" (p. 1820). We agree that an important next step in this area of research is a close examination of the distributional structure of tense marking in a school-based community sample of kindergarten children.

This Study

Evidence of the generalization of a bimodal distribution of tense marking to a kindergarten population remains to be established. In this study, tense marking was assessed with the Test of Early Grammatical Impairment (TEGI) Screening Test (Rice & Wexler, 2001) in a population-based sample of kindergarten speakers for whom no a priori classification or grouping criteria were applied. We hypothesized that if the bimodal distribution hypothesis was confirmed in the data, then it was expected that an empirically derived two-cluster distribution of TEGI Screening Test scores would result. If, on the other hand, analyses revealed either a single continuous structure or multiple cluster structures 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.

This study explored whether evidence for a bimodal distribution of tense marking exists in the 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 indicate the existence of a latent class of children who cluster around lower levels of accuracy and apart from a separate latent class of children with higher levels of accuracy?

Method

Targeted Kindergarten Population

In this study, cluster analysis, a conventional approach for testing categorical (or taxonic) boundaries, was run on data collected as part of a study of the grammatical skills of kindergarten children (Weiler, 2014). Specifically, this analysis focused on data collected in the fall of the 2014–2015 school year within one public school district in Middle Tennessee. The school district targeted for recruitment lies in a county that, according to data from the

2010 U.S. Census, is racially homogenous. The vast majority of county residents, that is, 94.8%, identify themselves as White only (U.S. Bureau of the Census, 2014b). The school district is composed of three elementary schools. The percentage of non-Hispanic White students at the three district elementary schools ranged from 92% to 97% (individually at 96.6%, 94.6%, and 91.9%; Tennessee Department of Education, 2014). Because the effects of dialect differences (e.g., AAE, 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 suspected of predominantly 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).

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 Tennessee state average of 24.4% and the national average of 29.3% (U.S. Bureau of the Census, 2014c). This figure also contrasts with the percentage of parents of children from the TEGI standardization sample who completed 4 or more years of college (15.5%–40.7% depending on age and language group—normal or disordered; Rice & Wexler, 2001). The economic and educational attainment status in this county was not considered a threat to the validity of this study. Despite the reported under-identification of language problems in children of lower socioeconomic 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- to high-SES backgrounds (Roy et al., 2014). Moreover, Rice et al. (1998) reported that maternal education level did not predict tense-marking growth over time among preschoolers with TL or early school–age children with SLI.

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. 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 to 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 this study are a dialectical feature of SWE speakers with unimpaired language skills. In their examination of grammatical features in the spontaneous language samples of nineteen 6-year-old speakers of a rural version of SWE with TL, Oetting and McDonald (2001) reported infrequent omissions of PT and 3S markers. By contrast, the overall omissions of obligatory PT and 3S markers from fifteen 6-year-old SWE speakers with SLI from the same study were 1.7–6.7 times greater than those of their SWE-speaking peers with TL, with the greatest difference occurring for PT markers. Cleveland and Oetting (2013) further quantified some of the findings from the work of Oetting and McDonald and reported a statistically significant difference in the mean percent obligatory 3S verbs marked for tense by SWE-speaking 6-year-olds with TL (93%) as compared to SWE-speaking 6-year-olds with SLI (71%; Cohen's d = 1.06). Accordingly, there is reason to predict that the distributional pattern of kindergarten 3S and PT tense marking in SWE speakers follows the same trend as that observed in MAE speakers. Therefore, the possible presence of SWE speakers in this study was determined to pose a very minimal threat to validity.

Participants

All kindergartners in each of the three elementary schools within one school district were invited to participate in a speech-language screening at the outset 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 parental consent to participate. The rate of returned consent across the three elementary schools ranged from 73% to 83% per school. Five consented kindergartners were withdrawn from the study because they failed to meet eligibility criteria (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 kindergartners analyzed, 81 (54.7%) were boys. The mean age of the sample at the time of screening was 5;8 (years;months; SD = 5 months; range: 4;11–6;10). All participants were assigned to general education kindergarten classrooms. Race or ethnicity was not collected on individual participants; however, observation indicated that 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 the following 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 an augmentative and alternative communication 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 monomorphemic words such as "bus" and "bed," the word-final phonemes used to mark 3S (e.g., "Every day he paints") and PT (e.g., "Yesterday she cleaned"). Three children were withdrawn because they failed the TEGI Phonological Probe. (c) The child did not demonstrate English proficiency. Given that the demographics of the three elementary schools in the participating district were overwhelmingly non-Hispanic White (see above), this last circumstance was not addressed formally through the use of a parent or teacher questionnaire. One child was withdrawn because the teacher reported anecdotally to the first author that the child was a native Spanish speaker with very limited English proficiency. This report was consistent with the examiner's (first author) verbal interaction with the child.

To ensure that students with potential linguistic vulnerabilities met the basic criterion for nonverbal cognitive functioning (i.e., NVIQ not consistent with classification of intellectual disability), the 51 children who failed to meet the TEGI manual–recommended criterion scores[2] for the TEGI Third Person Singular (3S) Probe, the TEGI Past Tense (PT) Probe, or the TEGI Screening Test 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 nine children scored in the low average range (standard score between 70 and 84). Given evidence of comparably compromised academic, social participation, and subjective well-being outcomes between children with language impairment who have at least average NVIQ and those who have low-average NVIQ (Tomblin, 2008), all 51 students were included in the analyses. Administration of the PTONI to all study participants was not feasible due to limited screening time available at the schools and a limited number of study team members.

Speech-Language Screening Battery

In the TEGI 3S Probe, the child was shown 11 pictures (one 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 3S subject to evaluate the child's production of the 3S tense marker in obligatory contexts (e.g., "She teaches"). Scoring of the TEGI 3S Probe was carried out following the guidelines delineated by Rice and Wexler (2001) in the TEGI manual. Child responses were scored correct for inclusion of the obligatory 3S marker and 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 reprompted with, for 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, the child was shown 20 pairs of pictures (two demonstrations, 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 PT in obligatory contexts (e.g., "He raked"). Scoring of the TEGI PT Probe was carried out following the guidelines delineated by Rice and Wexler (2001) in the TEGI manual. Similar to the scoring for 3S, child responses were scored correct for inclusion of the PT marker (i.e., correct regular or irregular PT form) and incorrect for omission of the PT marker (i.e., bare stem for regular or irregular verbs) when a subject is used (e.g., "He rake"). Additionally, 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

[2]These criterion score cut-points were developed by Rice and Wexler (2001) to reflect, at each 6-month age level between the ages 3;0 and 8;11, at least 80% sensitivity in separating the distribution of the group with language impairment in the standardization sample from the normal group of their standardization sample. According to Rice and Wexler (2001), "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 group designs discussed above, it was not expected that the TEGI manual–recommended cut-points would align exactly with any cluster boundaries found in the current study. Specifically, children forming the group with language impairment 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 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 standardization group with language impairment, 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).

a sentential subject were reprompted 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).

Procedure

The participants were administered individually (in the following order) a screening battery consisting of the TEGI Phonological Probe, the TEGI 3S Probe, and the TEGI PT Probe, as well as the Test of Articulation Performance–Screen (TAP-S; Bryant & Bryant, 1983). All child responses were audio-recorded with external microphones attached to a high-fidelity recorder. Child responses on the TEGI Phonological Probe items were marked correct or incorrect as to the inclusion of a nonmorphemic final consonant on the protocol form at the time of administration. Child responses on the TEGI 3S and PT Probe items were orthographically transcribed on protocol forms at the time of administration. Child responses on the TAP-S responses were transcribed phonetically on protocol forms to indicate any error responses. Children who failed to meet the author-recommended age-referenced 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 authors (certified speech-language pathologists [SLPs]) and 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 the authors prior to collecting data. All team members had experience in administration of single-word articulation tests. The assessment team for every child tested included a lab member who was a certified SLP with experience working in elementary schools. The PTONI was administered by the first author.

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 this 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” (e.g., full and individual evaluation in accordance with the Individuals with Disabilities Education Act; Rice & Wexler, 2001, p. 8).

The creation of the TEGI Screening Test score from the individual 3S and PT percent correct scores was supported psychometrically by reliability testing of the scores derived from this study. A high Cronbach’s coefficient alpha value of .907 was derived for the TEGI Screening Test score. 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 value 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 error 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 Checking and Reliability

Online 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, incorrect, or unscorable. Unscorable responses were rare, consisting of only 1.15% (17/1,480) of the total TEGI 3S items administered and 1.13% (2,634/2,664) of the total TEGI PT items administered. The majority of unscorable responses for both probes (3S: 76%; PT: 80%) were use of a verb form or tense other than the targeted form (e.g., “She was planting”). The team member who recorded the child’s responses online did not double-score that child’s responses. Scoring discrepancies were reviewed by a third team member and then were resolved by mutual consensus between the double scorer and the third team member via item-level comparison of score coding on the online protocol form. PTONI scoring was checked in the same manner. For reliability, a trained graduate research assistant retranscribed and 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. These steps were performed on blank protocol forms, and thus, the procedure was blinded to the original online scoring. Reliability was determined through comparisons between the initial checked scores and the reliability scores. 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. Therefore, reliability was achieved, and we proceeded with the initial checked scores.

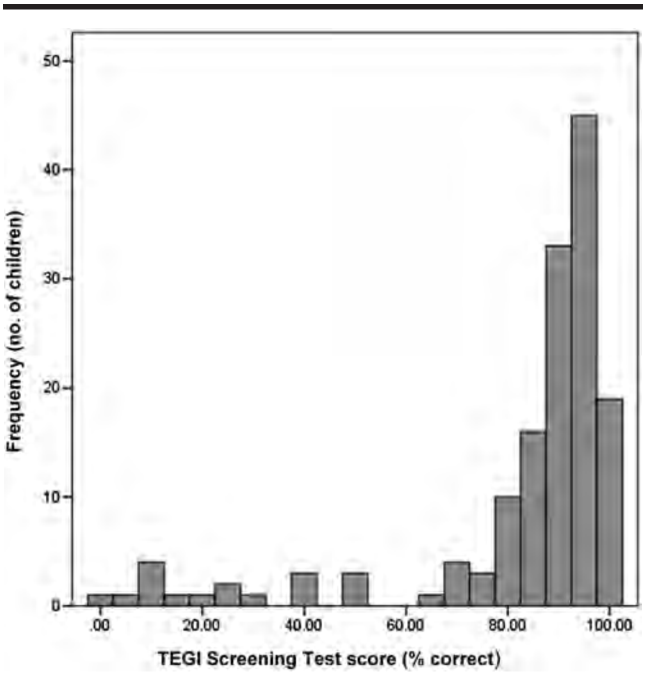
Results

The mean TEGI Screening Test score for the full sample of 148 children was 83.59 (SD = 23.13). The mean TEGI Screening Test score for the 34 children at or below the age criterion scores published in the TEGI manual was 50.22 (*SD* = 28.82). Of these 34 children, the mean age was 5;9 (*SD* = 6 months), and 53% (*n* = 18) were boys. The mean TEGI Screening Test score for the 114 children above the age criterion scores published in the TEGI manual was 93.23 (*SD* = 5.00). Of these 114 children, the mean age was 5;8 (*SD* = 5 months), and 55% (*n* = 63) were boys.

To analyze whether the TEGI Screening Test scores distributed non-normally, a Shapiro–Wilk test of normality was used. The histogram in Figure 2 illustrates the distribution. According to the Shapiro–Wilk test, the distribution of TEGI Screening Test scores deviated significantly from normality ($p < .001$) and had a substantial negative skew of −2.29. A distribution with a negative skew is desirable for a clinical marker where the majority of cases are expected to cluster around high levels of performance (Bishop, 2005; Dale et al., 2018).

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; IBM, n.d.). 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 (Hammett et al., 2003). The automatically generated best cluster solution is based on the Bayesian information criterion (BIC) for model selection among a finite set of models. The BIC is a measure of goodness-of-fit, with smaller values representing an increased fit (Mooi & Sarstedt, 2011). The model with the smallest BIC is preferred (Norusis, 2010). Although other information criteria, such as the Akaike information criterion (AIC), are available, the BIC was used in this study for several reasons. In a latent class analysis simulation study with a sample size ($n = 200$) comparable to that of this study, the BIC was a comparatively better indicator of the number of true classes than the AIC (Nylund et al., 2007). The BIC is also considered more parsimonious in that it seeks to identify a true model (i.e., the smallest correct model), whereas the AIC can choose an unnecessarily complex model (Dziak et al., 2020). Because we were interested in determining whether a model with two classes best fits the distributional structure of the TEGI Screening Test data, the BIC was deemed preferable.

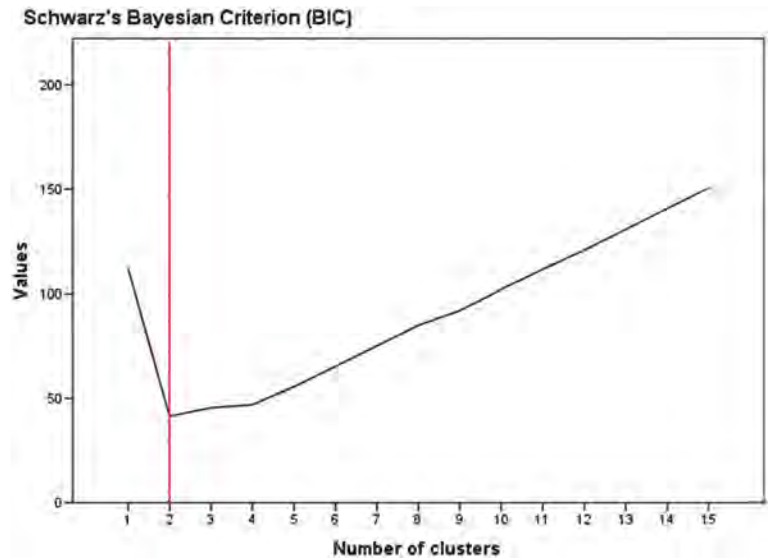
Figure 2. Distribution of Test of Early Grammatical Impairment (TEGI) Screening Test scores in a sample of kindergarten children from a single school district (*N* = 148; *M* = 83.59, *SD* = 23.13). Individual child test scores were calculated as a mean of the child's Third Person Singular Probe score and Past Tense Probe score (Rice & Wexler, 2001).



As evidenced by the vertical line in Figure 3, 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 Euclidean 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 clusters" 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, with values between .71 and 1.00 indicating a strong cluster structure. The two-cluster model solution described above resulted in an average silhouette value of .84 (*SD* = .12), indicating strong cluster structure with good cohesion within and separation across clusters. The average silhouette values of other possible cluster solutions were weaker than that of the two-cluster solution (three clusters: .78 [*SD* = .18]; four clusters: .67 [*SD* = .20]; five clusters: .67 [*SD* = .19]). Two-step cluster analysis was rerun in SPSS using the AIC instead of the BIC to determine the optimal cluster solution. The AIC-generated solution yielded the same two-cluster solution and membership profile as that described above using the BIC.

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 rerunning 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 data set). The results of both analyses were aligned highly with each other and with the original two-step

Figure 3. Cluster solutions and corresponding Bayesian information criterion (BIC) values. The vertical line illustrates the best fitting cluster model resulting from the two-step cluster analysis (IBM, n.d.).



analysis. In both of the half-samples, a two-cluster solution was found to best fit the data (see Figure 4). Moreover, the respective cluster sizes and individual cluster memberships were balanced across the two half-samples and, when aggregated, mirrored those from the full-sample two-step cluster analysis findings (see Table 1).

Additional confirmation of the best fitting cluster solution for these data was carried out through visual inspection of the dendrograms 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 dendrogram. 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 the "nearest neighbor" distance (Everitt et al., 2001). At each stage, a new candidate neighbor can be

Figure 4. Cluster solutions and corresponding Bayesian information criterion (BIC) values for the 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 right is for the second half of the random sample (n = 74; see Table 1). Vertical lines illustrate the best fitting cluster model resulting from the two-step cluster analysis (IBM, n.d.).

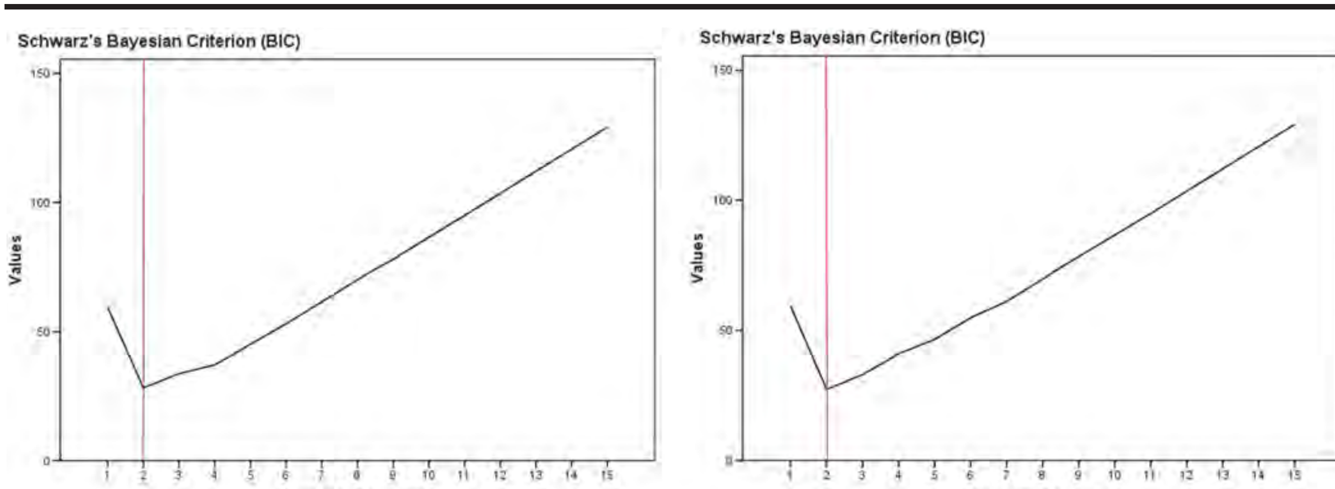


Table 1. Split-half reliability cluster sizes and members relative to the full sample (N = 148).

	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 (N = 148)		
Number of members	17	131
TEGI Screening Test,[a] M (SD)	25.12 (18.00)	91.18 (7.73)

Note. TEGI = Test of Early Grammatical Impairment (Rice & Wexler, 2001).

[a]Values represent percent correct scores.

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 dendrogram of the TEGI Screening Test scores for all 148 cases (or children) in Figure 5 visually illustrates this hierarchical clustering technique. The vertical height represents that distance at which each fusion is made (Everitt et al., 2001).

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 dendrogram is better carried out using Ward's method (Ward, 1963). Ward's hierarchical procedure maximizes between-clusters 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 dendrogram for the cluster number associated with the largest vertical-distance change in cluster fusion levels (Everitt et al., 2001). The dendrogram for the TEGI Screening Test score data illustrated in Figure 6 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. Moreover, the cluster sizes (ns = 131, 17) and cluster case members are identical to the results of the two-step method.

An additional check for the two-cluster solution was carried out following two numerical criteria for interpreting Ward's hierarchical cluster analysis offered by Lambert et al. (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). 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 SAS software (Version 9.4; SAS Institute, n.d.). 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 7 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 8 plots the cubic clustering criterion (CCC) statistic for estimating the number of clusters. Peaks in the

Figure 5. Single-linkage dendrogram for Test of Early Grammatical Impairment Screening Test scores resulting from the single-linkage hierarchical cluster analysis procedure (IBM, n.d.). Each numerical tick mark on the x-axis represents an individual case (N = 148). The y-axis values represent the rescaled distance (or dissimilarity) units where successive linkages of similar cases combine into clusters. Moving upward on the dendrogram, each joining of clusters is represented by a horizontal fusion of vertical lines. Toward the top of the dendrogram, the Low Cluster cases (n = 17) from the far-right side of the x-axis fuse with the remaining High Cluster cases (n = 131).

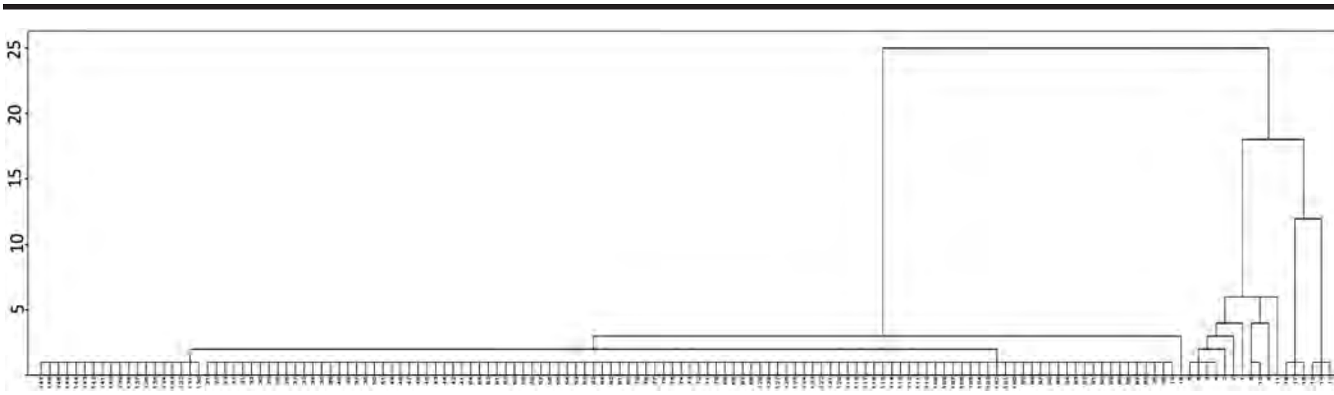
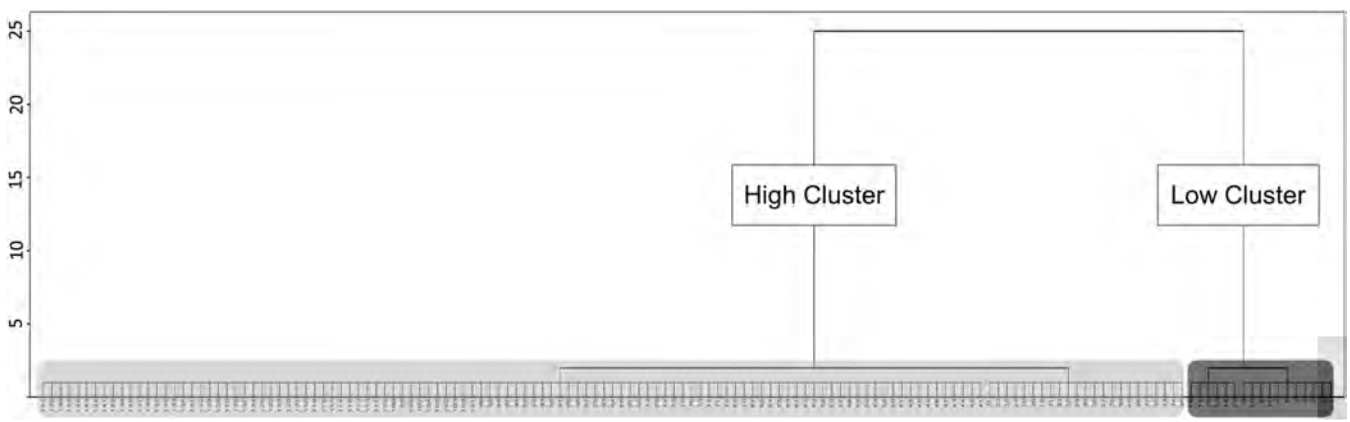


Figure 6. Ward linkage dendrogram for Test of Early Grammatical Impairment Screening Test scores created from Ward's hierarchical cluster analysis procedure (IBM, n.d.). Each numerical tick mark on the *x*-axis represents an individual case (N = 148). The *y*-axis values represent the rescaled distance (or dissimilarity) units where successive linkages of similar cases combine into clusters. Moving upward on the dendrogram, each joining of clusters is represented by a horizontal fusion of vertical lines. The cluster number associated with the largest vertical-distance change in cluster fusion levels is a solution rule (Everitt et al., 2001). The dendrogram shows that the two-cluster solution best satisfies this rule. Light-shaded cluster = High Cluster (n = 131); dark-shaded cluster = Low Cluster (n = 17).



plot of the CCC with values greater than 2 or 3 indicate good clusters; peaks with values between 0 and 2 indicate possible clusters (SAS Institute, n.d.). 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 value for two clusters 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 by the CCC.

A final check on the two-cluster solution was made using two measures, namely, C-index and point biserial correlation, which rank among the best indices for clustering partition quality (Milligan, 1981; Milligan & Cooper, 1985). The C-index (Hubert & Levin, 1976) compares the clustering partition obtained with the best partition theoretically possible given the number of clusters and the distances between all pairs of data points inside each cluster and in the entire data set (Desgraupes, 2013; Studer, 2013). C-index values range from 0 to 1; small values are indicative of good data partitioning (Studer, 2013). Using the mclust package (Scrucca et al., 2016) in R (R Core Team, 2017), the C-index value for the two-cluster solution found in this study was .006. The point biserial correlation (Milligan, 1980) is a measure indicating whether or not two corresponding data points are in the same cluster (Milligan & Cooper, 1985). Coefficient values range from −1 to 1; larger positive values reflect a better fit between the data and the obtained partition (Charrad et al., 2014). Using the NbClust package (Charrad et al., 2014) in R, the point biserial correlation for the two-cluster solution found in this study was

Figure 7. Difference between the actual R^2 value and the R^2 value expected by chance for each cluster solution (SAS Institute, n.d.). The two-cluster solution, having the greatest R^2 difference, met Lambert et al.'s (1998) criteria for a good solution.

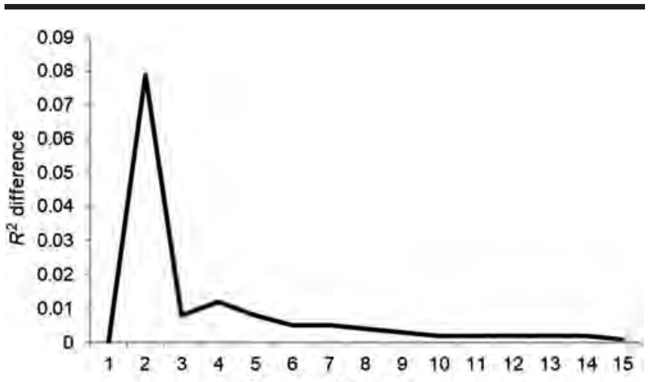
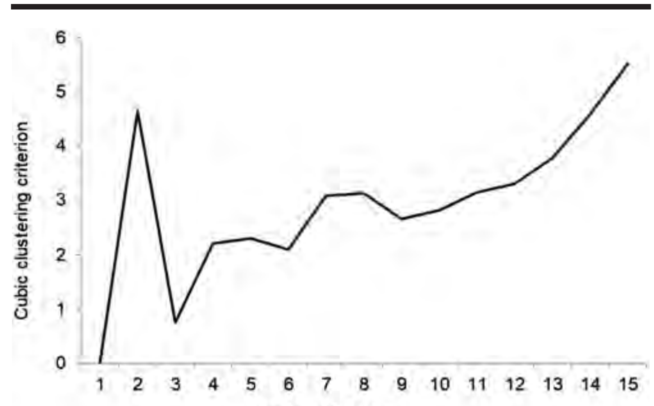


Figure 8. Cubic clustering criterion (CCC) statistic for estimating a cluster solution (Sarle, 1983; SAS Institute, n.d.). The local peak CCC value at two clusters met Lambert et al.'s (1998) criteria for an optimal cluster solution.



.90. By contrast, all point biserial correlation coefficients for other possible cluster solutions fell below .67.

Although not tied to the primary research questions, several correlational analyses were conducted to explore the relation between TEGI Screening Test scores and age and NVIQ across the two clusters (see Table 2). For the entire sample (N = 148), TEGI Screening Test scores did not significantly correlate with chronological age (p = .14). Nonsignificant correlations between age and TEGI Screening Test scores were similarly found for the individual clusters (Low Cluster: p = .21; High Cluster: p = .67). For the subset of 51 children administered the NVIQ measure (Low Cluster: n = 17; High Cluster: n = 34), TEGI Screening Test scores did not significantly correlate with PTONI standard scores (p = .99). Nonsignificant correlations between TEGI Screening Test scores and PTONI standard scores were similarly found for the individual clusters (Low Cluster: p = .63; High Cluster: p = .72).

Discussion

The robustness of tense marking as a clinical marker for SLI comes, in part, from demonstration of a bimodal distribution of this skill in clinically referred samples. Evidence of generalization of a bimodal distribution of tense marking to a population-based sample is, to the best of our knowledge, absent in the research literature. The purpose of this study was to address this gap by asking (a) if composite tense-marking scores collected from a population-based sample of kindergarten children within a single school district distribute non-normally and (b) if composite tense-marking scores indicate the existence of a latent class of children who cluster around lower levels of accuracy and apart from a separate latent class of children with higher levels of accuracy. TEGI Screening Test scores were calculated from the TEGI PT and 3S Probes administered early in the school year. The non-normal distribution and two-cluster solution offer evidence in support of tense marking as a bimodally distributed skill in the kindergarten population. A categorical structure rather than a dimensional structure best fit these data. As reflected in Figure 9, 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, a 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 3 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 an identifiable boundary delineating typical from atypical tense marking in a population-based sample of kindergartners.

Correlational results offer evidence that the difference in tense marking across the two clusters is not likely attributable to child chronological age. Additionally, the absence of a correlation between TEGI Screening Test scores and PTONI standard scores for the subset of children administered the PTONI is suggestive that nonverbal intelligence did not drive the cluster separation; however, see the Limitations section below for interpretation. A cluster difference in single-word speech production accuracy, as measured by the TAP-S articulation screener, was noted (see Table 3). This finding is unsurprising given that the estimate of speech delay prevalence (9.91%) in kindergarten-age children with language impairment is over 2.5 times greater than the overall prevalence of speech delay in 6-year-old children (3.8%; Shriberg et al., 1999).

As illustrated in Figure 1, in Rice and Wexler (1996), the threshold value separating the two distributions appeared

Table 2. Correlations between TEGI Screening Test scores and child age and TEGI Screening Test scores and PTONI standard scores.

TEGI Screening Test score	Age (months)	PTONI[a] (SS)
TEGI Screening Test score		
Combined clusters	.12	.01
Low cluster	.32	.13
High cluster	−.04	−.07

Note. PTONI = Primary Test of Nonverbal Intelligence (Ehrler & McGhee, 2008); SS = standard score; TEGI = Test of Early Grammatical Impairment (Rice & Wexler, 2001).

[a]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 Third Person Singular Probe, the Past Tense Probe, or the TEGI Screening Test (see Footnote 2).

Figure 9. Distribution of individual Test of Early Grammatical Impairment (TEGI) Screening Test scores (N = 148; Rice & Wexler, 2001). Shaded circles = Low Cluster case (n = 17), open circles = High Cluster case (n = 131); dotted line = Low Cluster curve, solid line = High Cluster curve.

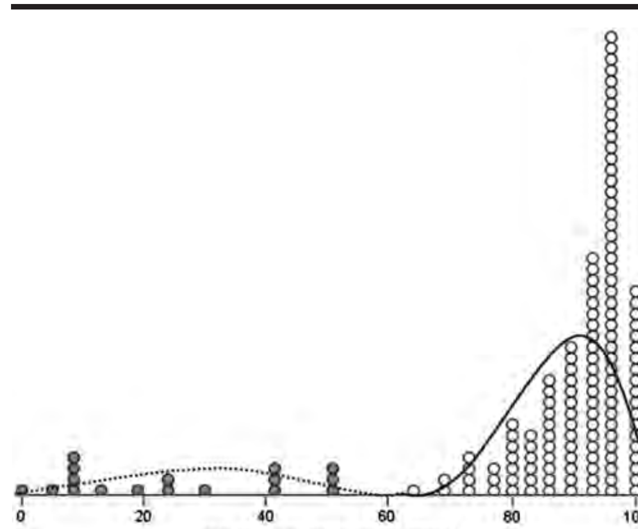


Table 3. Participant characteristics and testing summary for the total sample and by cluster.

Variable	Total sample *M* (*SD*)	Low Cluster[a] *M* (*SD*)	Low Cluster[a] Range	High Cluster[b] *M* (*SD*)	High Cluster[b] Range
Age (month)	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[c]	95.85 (14.69)	96.35 (16.20)	70–125	95.59 (14.12)	74–139
TAP-S AQ	95.69 (17.71)	78.06 (14.65)	< 58–109	97.98 (16.80)	< 57–118

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

[a]N = 17 (seven 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 Third Person Singular Probe, the Past Tense Probe, or the TEGI Screening Test (see Footnote 2).

to reach maximum discriminant accuracy in separating 5-year-olds with SLI from age-matched peers with TL at a score cut-point of approximately 65%–75% (Rice, 1998). Visual inspection of the two-cluster distribution in this study suggests that a broadly similar threshold value of approximately 60% separates the two clusters (see Figure 9). Children in the Low Cluster comprise 11.5% (17/148) of the study sample (NVIQ = 70–84: 2.0%; NVIQ ≥ 85: 9.5%). Although confirmatory diagnostic testing was not a part of this study, this percentage value of 11.5% is consistent with DLD prevalence estimates of 7%–13% (Beitchman et al., 1986; Norbury et al., 2016; Rice, 2020). Additionally, the percentage value of 9.5% of Low Cluster children with at least average NVIQ (i.e., ≥ 85) may be taken as a general prevalence estimate of those children at risk for SLI in the study sample. Comparisons to established SLI prevalence rates of 7%–8% reveal that the membership size of the Low Cluster with at least average NVIQ is broadly consistent with prevalence expectations. 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 gender imbalance is consistent with the 1.43:1 ratio of boys to girls found in the Low Cluster as compared to a 1.18:1 ratio in the High Cluster of the present investigation.

Findings from this study offer an important step forward in elucidating the status of tense-marking proficiency in a population-based sample of kindergartners. Central to the bimodal distribution hypothesis of tense marking among kindergarten children is the presence of an identifiable boundary separating a category of children who cluster at the upper end of the distribution from a category of children who cluster at the bottom of the distribution (Rice, 2000). We expect, in a population-based study, such a boundary to capture a broader swath of children with or at risk for language impairment associated with diagnoses beyond SLI. Tense-marking deficits, relative to the chronological age expectations of peers with TL, have been also documented among kindergarten children with nonspecific language impairment (i.e., NVIQ: −2 to −1 *SD*; Rice et al., 2004) who would fall under the DLD umbrella, preschool and kindergarten-age children with hearing loss (e.g., Guo et al., 2013; Werfel, 2018), young elementary-age children with Down syndrome (e.g., Eadie et al., 2002), school-age children with autism spectrum disorder and language impairment (e.g., Roberts et al., 2004), and school-age boys with fragile X syndrome (Sterling et al., 2012). As such, a bimodal distribution of kindergarten tense marking in a community sample can be taken as a valid indicator of language impairment presence and/or risk across a range of etiologies.

Limitations

Although a bimodal distribution of kindergarten tense marking was indicated in this study, it cannot be assumed that this finding is, in and of itself, diagnostically meaningful. Therefore, future studies that include confirmatory diagnostic accuracy testing with a valid and reliable reference standard for language impairment are recommended. Such studies would benefit additionally from methodological considerations that address some of the other limitations of this study.

Even though parental consent for participation 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% in Tomblin et al., 1997; 78% in Oetting, 2014).

The criteria for SLI (as well as DLD) stipulate that the deficit in language ability cannot be attributed to hearing loss (Bishop, 2017; Leonard, 2014). None of the 148 children tested had a visually apparent hearing aid or cochlear implant, nor did any children display behaviors during testing to suggest that they could not adequately hear the examiner. Still, it cannot be ruled out that some of the children 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-age population ranging from 3.1% (Mehra et al., 2009) to 5.2% (Bess et al., 1998), it is likely that a handful of children who participated in this study may have been excluded had hearing status been assessed.

The presence of low nonverbal intelligence is another exclusionary criterion for SLI and DLD, although cut-points in the literature range from < 85 to < 70 (Bishop, 2017; Gallinat & Spaulding, 2014; Leonard, 2014). In this study, nonverbal intelligence status was established by administering the PTONI to approximately one-third of the sample who failed to meet TEGI manual–recommended age cutoff scores for the TEGI 3S Probe, the TEGI PT Probe, or the TEGI Screening Test (3S + PT). The range of PTONI standard scores (70–139) in this subset of participants suggests a limited likelihood of intellectual disability (i.e., NVIQ < 70) in the population studied. As discussed in the Method section, the rationale for partial PTONI administration was based on resource limitations related to available research team personnel as well as limited screening time available in the schools. Ideally, all participants would have been administered the PTONI to avoid verification bias (Dollaghan, 2007). As it stands, however, the NVIQ status of the majority of participants in the High Cluster (n = 97) was not determined through direct testing. The nonsignificant cluster comparison on PTONI scores, therefore, should be viewed cautiously and tentatively because it did not factor in those children from the High Cluster who were not administered this measure. We cannot assume that the mean PTONI performance of the select High Cluster children administered this measure would have generalized to all High Cluster children. Future studies should be designed such that all participating children are administered a measure of NVIQ. Only by doing so can verification bias be avoided and cluster comparisons on NVIQ be made with confidence.

Recall that children were administered the PTONI if they failed the TEGI 3S Probe, the TEGI PT Probe, or the TEGI Screening Test. However, only the TEGI Screening Test scores (average of 3S + PT) were included in the cluster analysis. The finding that a number of children (n = 34) who fell below the TEGI manual–recommended age criterion score for the 3S Probe, the PT Probe, or the TEGI Screening Test score (average of 3S + PT) but, nonetheless, were assigned membership to the low tail of the High Cluster is noteworthy but not altogether surprising. Of these 34 children, half (n = 17) scored above criterion on the TEGI Screening Test but failed either the TEGI 3S (n = 3) or the TEGI PT (n = 14). Notably, standardization groups used for test development may not be entirely representative of local populations (Spaulding et al., 2006). Indeed, we would not expect perfect correspondence. As noted in the Method section, the TEGI standardization sample came from households where parents were more educated than adults in the county where this study took place. Additionally, as described in Footnote 2, there are several factors related to the recruitment of children for the Language Disorder Group of the TEGI standardization sample that could contribute to TEGI Screening Test age criterion scores that are higher than the boundary between the clusters observed in this study. Children in the Language Disorder Group of the TEGI standardization sample may have been included based on language testing up to 15 months old (inviting the possibility of the resolution or reduction of language difficulties) or due to language difficulties in areas outside of grammatical morphology (e.g., vocabulary) on omnibus tests of language. Although the bimodal distribution of tense marking held up in the present sample, factors such as these may have contributed to the difference in boundaries between groups (i.e., normal, disordered) from the TEGI standardization sample and between clusters in the population of this study.

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 the production accuracy of the entire word. In other words, if any sound within a word is produced in error, the item is scored 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 draw conclusions about the exact speech status (e.g., typical, delayed, developmentally appropriate errors) of the participants. Future studies in this area should include speech production measures that capture phoneme-level accuracy to better explain the possible linguistic interplay between grammatical tense marking and articulation skill.

The results of this study are limited by the unverified dialect status of the participants. The overwhelmingly non-Hispanic White demographic composition of the elementary schools the participants attended and of the county in which they resided is highly consistent with suspected MAE dialect. However, no formal measures of dialect were included in this study. Therefore, it remains possible that non-MAE dialect features could have been present in some of the participants, which may have, in turn, impacted our findings. Despite this possibility, we are reasonably confident that the overall results were not appreciably impacted by non-MAE dialect features associated with race or ethnicity. For example, the race of all 17 children in the Low Cluster was estimated, based on observation, as non-Hispanic White. The three total participants with an estimated observed African American race fell within the High Cluster (TEGI Screening Test scores: 73.5, 91.5, and 86.5). Additionally, as described in the Method, although findings by

Oetting and colleagues (Cleveland & Oetting, 2013; Oetting & McDonald, 2001) indicate that SWE dialect should not impact TEGI Screening Test responses in such a way as to have influenced the results, we cannot rule out this possibility entirely (Cleveland & Oetting, 2013; Oetting & McDonald, 2001). Future studies in this area should therefore include formal dialect measures such as blind listener judgments and language sample analysis.

It is presently unclear how our findings of a bimodal distribution of tense in the kindergarten sample assessed might generalize to other samples where dialect is clearly specified. If generalization were to be documented, then we might expect the cut-point separating the two clusters to shift depending upon the presence and density of dialect-specific nonmainstream zero forms of tense marking. Such studies would benefit from the inclusion of measures such as language sample analysis, blind listener judgments, and/or the Diagnostic Evaluation of Language Variation (Seymour et al., 2005) so that dialect-specific features of individual participants can be systematically coded for (see Oetting & McDonald, 2001; Washington & Craig, 1994) and factored into analyses as indicated.

Finally, the sample size (N = 148) of this study warrants consideration. On the one hand, the sample size met guidance for the adequacy for cluster analysis with a single variable offered by Dolnicar et al. (2014). Using results from a simulation study, these researchers found that a sample size of 70 times the number of variables proved adequate in yielding the correct solution for a known cluster structure. Although other research studies employing cluster analysis published in American Speech-Language-Hearing Association journals have had larger sample sizes than this study, they also included more variables. For example, Conti-Ramsden et al. (1997) ran *k*-means cluster analysis on 247 participants using six variables, whereas Tyler et al. (2008) carried out two-step cluster analysis on 153 participants using 51 variables.

On the other hand, whereas the sample size of this study permitted the detection of an apparent bimodal, two-cluster distribution of kindergarten tense-marking skill, the principles of the central limit theorem suggest that such detection may be limited in larger sample sizes. According to the central limit theorem, the distribution sampling mean approximates normality if the sample size is large enough even if the population sample itself is not normal (e.g., skewed, bimodal, multimodal). If the population distribution is normal, then sample sizes of 30 are typically sufficient for the central limit theorem to hold true. For population distributions that are not normally distributed, such as kindergarten tense marking, larger sample sizes (i.e., $N > 30$) may be required for the central limit theorem to take hold (Spiegel & Stephens, 1999). As such, it is possible that the sample size in this study facilitated the detection of an underlying bimodal distribution that might be otherwise undetectable in larger samples where normalization under the central limit theorem is realized. A critical next step in this area of empirical inquiry, then, will be the application of cluster analysis to larger population-based cohorts to determine if a bimodal distribution structure of TEGI Screening Test scores remains detectable with larger sample sizes. If not, then the establishment of a cutoff criterion score for risk of language impairment may present challenges.

Clinical Implications and Future Directions

This study provides another layer of evidence in support of the clinical marker utility of assessing tense marking in young school-age children. Our replication and extension of Rice's bimodal distribution of kindergarten tense marking to a 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 should be considered 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.

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 replicates in the same school district across school years or in other populations that differ on factors such as geography, SES, dialect, age, and residential strata (e.g., urban, suburban, rural). If a bimodal distribution of tense marking indeed holds up across 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.

Future studies are needed specifically to learn more about the distributions of PT and 3S in kindergartners with and without SLI across English dialects. It will be critical for this work to include controls with TL who speak the same dialect as the participants with SLI. Oetting and colleagues (2019, 2016) have been trailblazers in related studies pertaining 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 et al., 2016). Using a dialect-informed probe of PT marking and applying empirically derived cut scores, this lab also reported diagnostic accuracies for SLI of 77% in AAE-speaking kindergartners and 86% in SWE-speaking kindergartners (Oetting et al., 2019). Notably, the optimal AAE cut score of 54% overt PT marker inclusion was lower than the SWE cut score of 73% overt PT marker inclusion, indicating distinct distributional structures across dialects. Finally, in addition to continuing efforts to characterize the distributional properties of tense marking, child language researchers are encouraged to empirically evaluate the distributional structure of other clinical markers for pediatric SLI (e.g., nonword repetition, sentence recall). As seen in this study, cluster analysis offers one possible approach for identifying subgroups based on performance on these and other measures, for example, oral reading tense marking in older elementary children (e.g., Werfel et al., 2017).

Looking further ahead, if population-based studies with different samples and concurrent measures of diagnostic accuracy confirm the presence and discriminant validity of a bimodal distribution of kindergarten tense-marking skill, then the use of the TEGI Screening Test as a universal screener may be indicated. Universal language screening at school entry carries significant potential for improving identification of children at risk for language impairment and may be conceptualized as part of the response to intervention or multi tiered system of supports framework used in many school districts (Adlof & Hogan, 2019; Redmond et al., 2019; Rice, 2020). Although universal screening for reading and math skills is commonplace in public schools in the United States (e.g., Fuchs et al., 2012), the same cannot be said for language screening, which is not mandated by most school districts (Christopulos & Kean, 2020). Whereas large-scale screening for oral language was once routine practice for school SLPs (Ehren & Nelson, 2005), it is currently rare (Hendricks et al., 2019). Additionally, whereas several research studies of speech-language screening were published in the 1980s (e.g., Blaxley et al., 1983; Culatta et al., 1983; Illerbrun et al., 1985), it has only been in the past few years that investigations of universal language screening of early elementary–age children at school (e.g., Hendricks et al., 2019; Weiler et al., 2018) and at pediatric well-child visits (Ebert et al., 2020) have resurfaced.

In summary, results of this study offer converging evidence of a previously documented bimodal distribution of tense-marking proficiency in kindergarten-age children. Rice (1998) documented such a distribution in a clinically ascertained sample using histogram inspection. Findings from our investigation further document a bimodal distribution of tense in a population-based sample using taxometric (cluster) analysis. The divergence between these findings and those from the work of Rudolph et al. (2019), who reported a unimodal distribution of tense marking, could be attributable to methodological differences related to child language task (conversation with a family member; TEGI-elicited production); differences in determination of the distributional structure (visual; cluster analysis); or, as discussed above relative to the central limit theorem, differences in sample size. Demonstration of a bimodal distribution of tense marking in future studies with carefully defined population samples could strengthen the clinical marker evidence and utility of this linguistic feature.

Acknowledgments

This work was supported by U.S. Department of Education "Preparation of Leadership Personnel" Grants H325D080075 and H325K140087, awarded to C. Melanie Schuele, as well as by the American Speech-Language-Hearing Foundation Student Research Grant in Early Childhood Language Development and the Vanderbilt Institute for Clinical and Translational Research Voucher Award, both awarded to Brian Weiler.

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*(2), 360–372. https://doi.org/10.1044/2015_JSLHR-L-14-0003

Adlof, S. M., & Hogan, T. P. (2019). If we don't look, we won't see: Measuring language development to inform literacy instruction. *Policy Insights from the Behavioral and Brain Sciences, 6*(2), 210–217. https://doi.org/10.1177/2372732219839075

Aldenderfer, M. S., & Blashfield, R. K. (1984). Cluster analysis. *Sage*. https://doi.org/10.4135/9781412983648

Appalachian Regional Commission. (n.d.). *Counties in Appalachia*. https://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*(4), 148–158. https://doi.org/10.1044/lle21.4.148

Beitchman, J. H., Nair, R., Clegg, M., & Patel, P. G. (1986). Prevalence of speech and language disorders in 5-year-old kindergarten children in the Ottawa-Carleton region. *Journal of Speech and Hearing Disorders, 51*(2), 98–110. https://doi.org/10.1044/jshd.5102.98

Beitchman, J. H., Wilson, B., Johnson, C. J., Atkinson, L., Young, A., Adlaf, E., Escobar, M., & Douglas, L. (2001). Fourteen-year follow-up of speech/language-impaired and control children: Psychiatric outcome. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(1), 75–82. https://doi.org/10.1097/00004583-200101000-00019

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*(5), 339–354. https://doi.org/10.1097/00003446-199810000-00001

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). Erlbaum.

Bishop, D. V. M. (2005). DeFries–Fulker analysis of twin data with skewed distributions: Cautions and recommendations from a study of children's use of verb inflections. *Behavior Genetics, 35*(4), 479–490. https://doi.org/10.1007/s10519-004-1834-7

Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language & Communication Disorders, 49*(4), 381–415. https://doi.org/10.1111/1460-6984.12101

Bishop, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders, 52*(6), 671–680. https://doi.org/10.1111/1460-6984.12335

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*(5), 600–615. https://doi.org/10.1080/13682820802259662

Blaxley, L., Clinker, M., & Warr-Leeper, G. (1983). Two language screening tests compared with developmental sentence scoring.

Language, Speech, and Hearing Services in Schools, 14(1), 38–46. https://doi.org/10.1044/0161-1461.1401.38

Brownlie, E. B., Bao, L., & Beitchman, J. (2016). Childhood language disorder and social anxiety in early adulthood. *Journal of Abnormal Child Psychology, 44*(6), 1061–1070. https://doi.org/10.1007/s10802-015-0097-5

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

Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). NbClust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software, 61*(6), 1–36. https://doi.org/10.18637/jss.v061.i06

Christopulos, T., & Kean, J. (2020). General education teachers' contribution to the identification of children with language disorders. *Perspectives of the ASHA Special Interest Groups, 5*(4), 770–777. https://doi.org/10.1044/2020_PERSP-19-00166

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*(4), 604–614. https://doi.org/10.1044/1058-0360(2013/12-0122)

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

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

Conti-Ramsden, G., Crutchley, A., & Botting, N. (1997). The extent to which psychometric tests differentiate subgroups of children with SLI. *Journal of Speech, Language, and Hearing Research, 40*(4), 765–777. https://doi.org/10.1044/jslhr.4004.765

Conti-Ramsden, G., & Durkin, K. (2012). Postschool educational and employment experiences of young people with specific language impairment. *Language, Speech, and Hearing Services in Schools, 43*(4), 507–520. https://doi.org/10.1044/0161-1461(2012/11-0067)

Conti-Ramsden, G., Durkin, K., Simkin, Z., & Knox, E. (2009). 1Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *International Journal of Language & Communication Disorders, 44*(1), 15–35. https://doi.org/10.1080/13682820801921601

Conti-Ramsden, G., Mok, P., Durkin, K., Pickles, A., Toseeb, U., & Botting, N. (2019). Do emotional difficulties and peer problems occur together from childhood to adolescence? The case of children with a history of developmental language disorder (DLD). *European Child & Adolescent Psychiatry, 28*(7), 993–1004. https://doi.org/10.1007/s00787-018-1261-6

Culatta, B., Page, J. L., & Ellis, J. (1983). Story retelling as a communicative performance screening tool. *Language, Speech, and Hearing Services in Schools, 14*(2), 66–74. https://doi.org/10.1044/0161-1461.1402.66

Dale, P. S., Rice, M. L., Rimfeld, K., & Hayiou-Thomas, M. E. (2018). Grammar clinical marker yields substantial heritability for language impairments in 16-year-old twins. *Journal of Speech, Language, and Hearing Research, 61*(1), 66–78. https://doi.org/10.1044/2017_JSLHR-L-16-0364

Desgraupes, B. (2013). *Clustering indices.* University of Paris Ouest - Lab Modal'X.

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

Dolnicar, S., Grün, B., Leisch, F., & Schmidt, K. (2014). Required sample sizes for data-driven market segmentation analyses in tourism. *Journal of Travel Research, 53*(3), 296–306. https://doi.org/10.1177/0047287513496475

Dziak, J. J., Coffman, D. L., Lanza, S. T., Li, R., & Jermiin, L. S. (2020). Sensitivity and specificity of information criteria. *Briefings in Bioinformatics, 21*(2), 553–565. https://doi.org/10.1093/bib/bbz016

Eadie, P. A., Fey, M. E., Douglas, J. M., & Parsons, C. L. (2002). Profiles of grammatical morphology and sentence imitation in children with specific language impairment and Down syndrome. *Journal of Speech, Language, and Hearing Research, 45*(4), 720–732. https://doi.org/10.1044/1092-4388(2002/058)

Ebert, K. D., Ochoa-Lubinoff, C., & Holmes, M. P. (2020). Screening school-age children for developmental language disorder in primary care. *International Journal of Speech-Language Pathology, 22*(2), 152–162. https://doi.org/10.1080/17549507.2019.1632931

Ehren, B. J., & Nelson, N. W. (2005). The responsiveness to intervention approach and language impairment. *Topics in Language Disorders, 25*(2), 120–131. https://doi.org/10.1097/00011363-200504000-00005

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

Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster analysis* (4th ed.). Arnold.

Fuchs, D., Fuchs, L., & Compton, D. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78*(3), 263–279. https://doi.org/10.1177/001440291207800301

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*(4), 1363–1382. https://doi.org/10.1044/2014_JSLHR-L-12-0363

Guo, L.-Y., Spencer, L. J., & Tomblin, J. B. (2013). Acquisition of tense marking in English-speaking children with cochlear implants: A longitudinal study. *Journal of Deaf Studies and Deaf Education, 18*(2), 187–205. https://doi.org/10.1093/deafed/ens069

Hammett, L. 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*(4), 442–468. https://doi.org/10.1598/RRQ.38.4.2

Hendricks, A. E., Adlof, S. M., Alonzo, C. N., Fox, A. B., & Hogan, T. P. (2019). Identifying children at risk for developmental language disorder using a brief, whole-classroom screen. *Journal of Speech, Language, and Hearing Research, 62*(4), 896–908. https://doi.org/10.1044/2018_JSLHR-L-18-0093

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*(4), 835–862. https://doi.org/10.1017/S0305000911000365

Hubert, L. J., & Levin, J. R. (1976). A general statistical framework for assessing categorical clustering in free recall. *Psychological Bulletin, 83*(6), 1072–1080. https://doi.org/10.1037/0033-2909.83.6.1072

IBM. (n.d.). *SPSS Statistics* (Version 23) [Computer software].

Illerbrun, D., Haines, L., & Greenough, P. (1985). Language identification screening test for kindergarten: A comparison with four screening and three diagnostic language tests. *Language, Speech, and Hearing Services in Schools, 16*(4), 280–292. https://doi.org/10.1044/0161-1461.1604.280

Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology, 19*(1), 51–66. https://doi.org/10.1044/1058-0360(2009/08-0083)

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., Brownlie, E. B., Douglas, L., Taback, N., Lam, I., & Wang, M. (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*(3), 744–760. https://doi.org/10.1044/jslhr.4203.744

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

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*(4), 1326–1340. https://doi.org/10.1044/2015_JSLHR-L-14-0348

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*(1), 47–57. https://doi.org/10.1016/S0149-7189(97)00044-X

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*(6), 1401–1416. https://doi.org/10.1044/1092-4388(2009/08-0142)

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). MIT Press. https://doi.org/10.7551/mitpress/9152.001.0001

McGregor, K. K., Goffman, L., Van Horne, A. O., Hogan, T. P., & Finestack, L. H. (2020). Developmental language disorder: Applications for advocacy, research, and clinical service. *Perspectives of the ASHA Special Interest Groups, 5*(1), 38–46. https://doi.org/10.1044/2019_PERSP-19-00083

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*(4), 461–472. https://doi.org/10.1016/j.otohns.2008.12.022

Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika, 45*(3), 325–342. https://doi.org/10.1007/BF02293907

Milligan, G. W. (1981). A Monte Carlo study of thirty internal criterion measures for cluster analysis. *Psychometrika, 46*(2), 187–199. https://doi.org/10.1007/BF02293899

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*(2), 159–179. https://doi.org/10.1007/BF02294245

Mooi, E., & Sarstedt, M. (2011). Cluster analysis. In E. Mooi & M. Sarstedt (Eds.), *A concise guide to market research* (pp. 237–284). Springer-Verlag.

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], The University of Iowa.

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *The Journal of Child Psychology and Psychiatry, 57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Norusis, M. J. (2010). *PASW statistics 18: Statistical procedures companion*. Prentice Hall.

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

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Oetting, J. B. (2014, November). *Language impairment in children who speak nonmainstream dialects* [Paper presentation]. American Speech-Language-Hearing Association's 24th Annual Research Symposium, Orlando, FL, United States. https://doi.org/10.1044/cred-pvd-c14001

Oetting, J. B., Berry, J. R., Gregory, K. D., Rivière, A. M., & McDonald, J. (2019). Specific language impairment in African American English and Southern White English: Measures of tense and agreement with dialect-informed probes and strategic scoring. *Journal of Speech, Language, and Hearing Research, 62*(9), 3443–3461. https://doi.org/10.1044/2019_JSLHR-L-19-0089

Oetting, J. B., & McDonald, J. L. (2001). Nonmainstream dialect use and specific language impairment. *Journal of Speech, Language, and Hearing Research, 44*(1), 207–223. https://doi.org/10.1044/1092-4388(2001/018)

Oetting, J. B., McDonald, J. L., Seidel, C. M., & Hegarty, M. (2016). Sentence recall by children with SLI across two nonmainstream dialects of English. *Journal of Speech, Language, and Hearing Research, 59*(1), 183–194. https://doi.org/10.1044/2015_JSLHR-L-15-0036

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of clinical markers of specific language impairment in adults. *Journal of Speech, Language, and Hearing Research, 53*(2), 414–429. https://doi.org/10.1044/1092-4388(2009/08-0016)

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Redmond, S. M., Ash, A. C., Christopulos, T. T., & Pfaff, T. (2019). Diagnostic accuracy of sentence recall and past tense measures for identifying children's language impairments. *Journal of Speech, Language, and Hearing Research, 62*(7), 2438–2454. https://doi.org/10.1044/2019_JSLHR-L-18-0388

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. https://doi.org/10.1044/2014_LSHSS-14-0045

Reilly, S., Tomblin, J. B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. M., & Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language & Communication Disorders, 49*(4), 416–451. https://doi.org/10.1111/1460-6984.12102

Rice, M. L. (1998). In search of a grammatical marker of language impairment in children. *SIG 1 Perspectives on Language Learning and Education, 5*(1), 3–7. https://doi.org/10.1044/lle5.1.3

Rice, M. L. (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* (pp. 17–34). Taylor & Francis.

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). Erlbaum. https://doi.org/10.4324/9781410610881

Rice, M. L. (2020). Clinical lessons from studies of children with specific language impairment. *Perspectives of the ASHA Special Interest Groups, 5*(1), 12–29. https://doi.org/10.1044/2019_PERSP-19-00011

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

Hearing Research, 47(4), 816–834. https://doi.org/10.1044/1092-4388(2004/061)
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*(6), 1239–1257. https://doi.org/10.1044/jshr.3906.1239
Rice, M. L., & Wexler, K. (2001). *Rice/Wexler Test of Early Grammatical Impairment*. The University of Kansas. https://cldp.ku.edu/rice-wexler-tegi
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*(6), 1412–1431. https://doi.org/10.1044/jslhr.4106.1412
Roberts, J. A., Rice, M. L., & Tager-Flusberg, H. (2004). Tense marking in children with autism. *Applied Psycholinguistics, 25*(3), 429–448. https://doi.org/10.1017/S0142716404001201
Roy, P., Chiat, S., & Dodd, B. (2014). *Language and socioeconomic disadvantage: From research to practice*. City, University of London.
Rudolph, J. M., Dollaghan, C. A., & Crotteau, S. (2019). The finite verb morphology composite: Values from a community sample. *Journal of Speech, Language, and Hearing Research, 62*(6), 1813–1822. https://doi.org/10.1044/2019_JSLHR-L-18-0437
Ruscio, J., & Ruscio, A. M. (2004). A nontechnical introduction to the taxometric method. *Understanding Statistics, 3*(3), 151–194. https://doi.org/10.1207/s15328031us0303_2
Sarle, W. S. (1983). *Cubic clustering criterion* (SAS Technical Report A-108). SAS Institute.
SAS Institute. (n.d.). SAS (Version 9.4) [Computer software].
Schwartz, H. D., & Conture, E. G. (1988). Subgrouping young stutterers: Preliminary behavioral observations. *Journal of Speech and Hearing Research, 31*(1), 62–71. https://doi.org/10.1044/jshr.3101.62
Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal, 8*(1), 289–317. https://doi.org/10.32614/RJ-2016-021
Seymour, H. N., Bland-Stewart, L., & Green, L. J. (1998). Difference versus deficit in child African American English. *Language, Speech, and Hearing Services in Schools, 29*(2), 96–108. https://doi.org/10.1044/0161-1461.2902.96
Seymour, H. N., Roeper, T., De Villiers, J. G., & De Villiers, P. A. (2005). *Diagnostic Evaluation of Language Variation, Norm Referenced: DELV-NR*. Ventris Learning.
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*(6), 1461–1481. https://doi.org/10.1044/jslhr.4206.1461
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 Schools, 37*(1), 61–72. https://doi.org/10.1044/0161-1461(2006/007)
Spaulding, T. J., Szulga, M. S., & Figueroa, C. (2012). Using norm-referenced tests to determine severity of language impairment in children: Disconnect between U.S. policy makers and test developers. *Language, Speech, and Hearing Services in Schools, 43*(2), 176–190. https://doi.org/10.1044/0161-1461(2011/10-0103)
Spiegel, M. R., & Stephens, L. J. (1999). *Schaum's outline of theory and problems of statistics* (3rd ed.). McGraw-Hill.
Sterling, A. M., Rice, M. L., & Warren, S. F. (2012). Finiteness marking in boys with fragile X syndrome. *Journal of Speech, Language, and Hearing Research, 55*(6), 1704–1715. https://doi.org/10.1044/1092-4388(2012/10-0106)
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*(2), 407–418. https://doi.org/10.1044/jslhr.4102.407
Studer, M. (2013). *WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R (Working Paper 24)*. Swiss National Centre of Competence in Research LIVES.
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*(5), 1275–1278. https://doi.org/10.1044/jslhr.4205.1275
Tennessee Department of Education. (2014). *State report card*. https://www.tn.gov/education/topic/report-card
Tennessee Department of Education. (2018). *Speech or language impairment evaluation guidance*.
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* (pp. 93–114). Psychology Press.
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*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245
Tyler, R., Coelho, C., Tao, P., Ji, H., Noble, W., Gehringer, A., & Gogel, S. (2008). Identifying tinnitus subgroups with cluster analysis. *American Journal of Audiology, 17*(2), S176–S184. https://doi.org/10.1044/1059-0889(2008/07-0044)
U.S. Bureau of the Census. (2014a). *Current population survey, annual social and economic supplement; Small area income and poverty estimates*. Retrieved January 28, 2016, from https://www.census.gov/quickfacts/table/IPE120214/00,47085
U.S. Bureau of the Census. (2014b). *Population estimates program*. Retrieved January 21, 2016, from https://www.census.gov/quickfacts/table/IPE120214/00,47085,47#headnote-js-a
U.S. Bureau of the Census. (2014c). *2010–2014 American community survey, 5-year estimates*. Retrieved January 28, 2016, from https://www.census.gov/quickfacts/table/IPE120214/00,47085,47
Ward, J. H., J. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845
Washington, J., & Craig, H. (1994). Dialectal forms during discourse of poor, urban, African American preschoolers. *Journal of Speech and Hearing Research, 37*(4), 816–823. https://doi.org/10.1044/jshr.3704.816
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*(1), 82–89. https://doi.org/10.1044/0161-1461(2004/009)
Weiler, B. (2014). *Participle –ed: The role of argument structure and interpretation* [Unpublished raw data].
Weiler, B., Schuele, C. M., Feldman, J. I., & Krimm, H. (2018). A multiyear population-based study of kindergarten language screening failure rates using the Rice Wexler Test of Early Grammatical Impairment. *Language, Speech, and Hearing Services in Schools, 49*(2), 248–259. https://doi.org/10.1044/2017_LSHSS-17-0071

Werfel, K. L. (2018). Morphosyntax production of preschool children with hearing loss: An evaluation of the extended optional infinitive and surface accounts. *Journal of Speech, Language, and Hearing Research, 61*(9), 2313–2324. https://doi.org/10.1044/2018_JSLHR-L-17-0406

Werfel, K. L., Hendricks, A. E., & Schuele, C. M. (2017). The potential of past tense marking in oral reading as a clinical marker of specific language impairment in school-age children. *Journal of Speech, Language, and Hearing Research, 60*(12), 3561–3572. https://doi.org/10.1044/2017_JSLHR-L-17-0115

Wolfram, W., & Christian, D. (1976). *Appalachian speech*. 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*(4), 345–357. https://doi.org/10.1044/1058-0360.0904.345