LSHSS

Research Article

Differentiating Children With and Without Language Impairment Based on Grammaticality

Sarita L. Eisenberg[a] and Ling-Yu Guo[b]

Purpose: This study compared the diagnostic accuracy of a general grammaticality measure (i.e., percentage grammatical utterance; PGU) to 2 less comprehensive measures of grammaticality—a measure that excluded utterances without a subject and/or main verb (i.e., percentage sentence point; PSP) and a measure that looked only at verb tense errors (i.e., percentage verb tense usage; PVT)—in differentiating children with and without language impairment.
Method: Two groups of 3-year-olds, 17 with language impairment and 17 with typical language, participated in a picture description task. PGU, PSP, and PVT were computed. Receiver operating characteristic curve analyses were conducted to determine the best cutoff value for each measure.
Results: All 3 measures demonstrated a sensitivity of 100%. PGU showed a specificity of 88%, and both PSP and PVT showed a specificity of 82%. In addition, PGU showed a larger positive likelihood ratio than the other 2 measures.
Conclusion: PGU, PSP, and PVT were all sensitive to language impairment. However, PGU was less likely than PSP and PVT to misclassify children with typical language. The resultant diagnostic accuracy makes PGU an appropriate measure to use to screen for language impairment.

Key Words: assessment, language sampling, syntax, language disorders, children

Although many children outgrow early language delays, the likelihood of doing so decreases for children who do not catch up by age 3 (Rescorla & Schwartz, 1990; Thal & Katich, 1996). This makes age 3 a critical time for identifying language impairment (LI). For 5-year-old children, Rice, Wexler, and Hershberger (1998) suggested using reduced usage of tense marking as a clinical marker of LI. In addition, Tomblin, Records, and Zhang (1996) suggested using 1.25 *SD*s below the mean (i.e., –1.25 *SD*) on two or more composite scores of a norm-referenced test as a diagnostic standard for identifying LI at age 5. However, identification of LI at younger ages remains difficult (Leonard, 1998; Rescorla & Lee, 2001).

[a]Montclair State University, Montclair, NJ
[b]University at Buffalo–The State University of New York, Buffalo

Correspondence to Sarita Eisenberg:
eisenbergs@mail.montclair.edu

Editor: Marilyn Nippold
Associate Editor: Cheryl Scott

Received October 12, 2011
Revision received March 19, 2012
Accepted May 15, 2012
DOI: 10.1044/0161-1461(2012/11-0089)

Diagnosis of LI typically involves the use of norm-referenced standardized tests (McCauley, 2001; Paul & Norbury, 2012). In a survey by Huang, Hopkins, and Nippold (1997), 81% of speech-language pathologists (SLPs) reported using norm-referenced standardized tests for identifying LI. Tests are also used in research studies to classify children as typically developing or having an LI (e.g., Leonard et al., 2007; Tomblin et al., 1997). However, based on their comparison of measures from language sample analysis (LSA; e.g., mean length of utterance [MLU], percentage structural errors) and scores from norm-referenced standardized tests, Dunn, Flax, Sliwinski, and Aram (1996) concluded that quantitative LSA measures may be more sensitive to LI in young children than norm-referenced standardized tests.

MLU is the most commonly used LSA measure (Hux, Morris-Friehe, & Sanger, 1993; Loeb, Kinsler, & Bookbinder, 2000) and has been reported to distinguish between LI and typical language (TL) groups (Rice et al., 2010; Scott & Windsor, 2000). Eisenberg, Fersko, and Lundgren (2001), however, noted that an MLU below the –1 *SD* screening cutoff recommended by Miller and Chapman (1981) cannot be interpreted as evidence for LI because this MLU value is within the range of TL performance and has a specificity of only 80% based on data from Klee, Schaffer, May, Membrino, and Mougey (1989). This is consistent with more recent

data provided by Rice et al. (2010), which showed the mean MLU for children with LI to be at –1 *SD* for TL children. Setting a lower cutoff for MLU at –1.5 *SD* increases specificity to 90% but reduces sensitivity to 63% (based on data from Klee et al., 1989). These data may explain why SLPs feel that LSA measures are more useful for diagnosing moderate or severe LI than for diagnosing milder LI (Hux et al., 1993).

Because children with LI have considerable difficulty using grammatical morphemes, with morphemes marking tense being particularly problematic, usage of tense marking has been proposed as a diagnostic marker for LI (Rice et al., 1998; Tager-Flusberg & Cooper, 1999). Rice and Wexler (2001) developed a standardized test, the Rice-Wexler Test of Early Grammatical Impairment (TEGI), which uses a composite verb tense measure to identify children with LI. The composite verb tense measure averages the scores from elicited production probes for third person singular *–s,* regular and irregular past tense, copula and auxiliary *be,* and auxiliary *do*. Although Rice et al. (1998) suggested using a composite tense marking measure as a clinical marker for children ages 4;6 (years;months) and older, the TEGI also provides performance data for younger children. For older 3-year-olds (3;6–3;11), the composite verb tense score of .62, which yielded good sensitivity (≥90%), showed only fair specificity (83%) based on the criteria of Plante and Vance (1994). Similarly, a score of .55 that yielded good specificity for this age showed only fair sensitivity (83%). For younger 3-year-olds, a composite verb tense score of .53 that yielded good sensitivity showed unacceptably low specificity (74%) based on the criteria of Plante and Vance (1994), whereas a score of .32 with good specificity showed unacceptably low sensitivity (70%; Rice & Wexler, 2001).

Bedore and Leonard (1998) developed a tense marking measure based on conversational samples, called the finite verb morphology composite (FVMC), to differentiate 38 children with and without LI between the ages of 3;7 and 5;9. The FVMC computes children's percentage correct use of copula and auxiliary *be,* third person singular *–s,* and past tense *–ed* in conversation. Discriminant analyses indicated that the FVMC has a sensitivity of 84% and a specificity of 100%. Goffman and Leonard (2000) compared the FVMC from language samples of nine children with specific language impairment (SLI) and their age-matched TL peers. Although the data of the TL children were obtained cross-sectionally, each child with SLI was tested four times over 2 years. The age range of the children with SLI was ∼3;4 to 5;3 during the study. Except for one data point for one child, all children with SLI produced an FVMC at or below –1 *SD* as compared to the TL children. Taken together, these studies showed that a composite tense measure had moderate to good accuracy in identifying or differentiating children with and without SLI, especially when the composite verb tense was computed from conversational language samples.

Although tense errors are potentially a diagnostic marker of LI, they are not the only difficulties that children with LI have (Dunn et al., 1996). Other difficulties that have been observed in children with LI may include, but are not limited to, argument structure errors (Grela & Leonard, 1997), pronoun errors (Loeb & Leonard, 1991; Moore, 2001), errors with grammatical morphemes other than tense markers (Leonard, Eyer, Bedore, & Grela, 1997; Watkins & Rice, 1991), and overreliance on general all-purpose verbs (Rice & Bode, 1993). Thus, if we use a composite verb tense measure to identify children with LI, it is possible that we may miss some children with LI who are relatively less affected in tense marking but more affected in other aspects of grammar. This suggests the importance of employing a more comprehensive measure for diagnosing children with LI.

One such measure is the sentence point that is part of the Developmental Sentence Scoring (DSS) analysis (Lee, 1974). The sentence point in the DSS analysis evaluates the grammaticality of 50 consecutive sentences in language samples based on the presence or absence of errors in grammatical morphemes, syntax, and/or semantics. A grammatical sentence receives a sentence point of 1; an ungrammatical sentence receives a sentence point of 0. Lee (1974) reported a mean sentence point score of 35.28 for forty 3-year-olds, corresponding to a 70% (i.e., 35.28/50 × 100%) grammaticality rate for the sentences included in the DSS analysis. However, because Lee reported only the mean without reporting variability, the reported sentence point score is not useful for clinical decision making. In addition, because the DSS excludes sentences without a main verb, including copula omissions (e.g., *He happy*), and nonimperative utterances without a subject (e.g., *Want a cookie*), Lee may have overestimated the overall rate of grammaticality.

To address the concern of overestimating children's level of grammaticality on the DSS analysis, Eisenberg, Guo, and Germezi (2012) proposed an alternative measure, percentage grammatical utterances (PGU), to investigate grammatical development in twenty 3-year-olds. Like the DSS analysis, PGU takes into account aspects of grammatical morphology and syntax as well as semantics (e.g., vocabulary usage). However, PGU includes utterances with copula omissions and nonimperative utterances without a subject in the computation. Study results indicated that the mean PGU in typical 3-year-olds was 71% (*SD* = 10%, range = 47%–89%), which was comparable to the 70% grammaticality level derived from the DSS sentence point score previously reported by Lee (1974).

The most common error that the children in the Eisenberg et al. (2012) study produced involved tense marking, but these errors accounted for only ∼30% of the total errors. The children in the study also produced errors on grammatical morphemes other than tense markers (27%), pronominal forms (13%), and argument structures (11%), as well as on the use of fragments (e.g., phrases without verbs, 3%). Another

16% of errors were categorized as *other,* which included lexical errors. The error analysis reiterated the importance of employing a more comprehensive measure to characterize young children's grammatical development. However, given the variability of PGU in 3-year-olds with TL, the accuracy of using PGU to differentiate children with and without LI as compared to a similar measure that excludes fragments (i.e., the DSS sentence point) or a more specific measure (e.g., tense composite) remains an open question.

The current study investigated the diagnostic accuracy of PGU, the DSS sentence point, and a tense composite in differentiating 3-year-olds with and without LI. We focused on 3-year-old children for several reasons. Children of this age become eligible for assessments by their local school district, and normative comparisons are required to qualify children for preschool services. In addition, children of this age can be difficult to test, so quantitative LSA measures, such as PGU, would be particularly useful at this age. We asked the following questions:

- How accurately does overall level of grammaticality, as measured by PGU, differentiate between LI and TL at age 3?
- Is PGU more accurate in differentiating between LI and TL than the DSS sentence point, which excludes sentences without a subject and/or main verb from the analysis?
- Is PGU more accurate in differentiating between LI and TL than a composite verb tense measure?

METHOD

Participants

Participants included 34 children between the ages of 3;0 and 3;11 who were recruited through nursery school programs, pediatricians, and SLPs in suburban New Jersey as well as through online announcements. Approval for this research was granted by the Montclair State University Institutional Review Board.

Seventeen of the children (7 girls; 10 boys) were classified into the LI group. Eleven of these children had been previously diagnosed with LI at age 2 (the LI2 group) and were referred to the study by the SLP from whom they were currently receiving treatment for language. The remaining six children were classified into the LI group at age 3 (the LI3 group) as part of the current study. All six children in the LI3 group were subsequently evaluated by an SLP and enrolled in treatment for language. We excluded children with more severe language deficits who were nonverbal or who were producing only single words and two-word combinations because we wanted children to be at least able to produce sentences so that we would have sufficient contexts for computing the target measures.

The criteria for diagnosing the children in the LI3 group was based on the suggestion by Paul and Norbury (2012) that diagnosis of LI must consider both environmental and norm-referenced expectations. A consideration of environmental expectations means that a child must demonstrate "a deficit that is big enough to be noticed by ordinary people such as parents and teachers" (p. 9). To meet this criterion, we used a parent rating scale adapted from Hadley and Rice (1993). The parent rating scale was a 14-item questionnaire that required the parent to rate aspects of the child's language development (e.g., ability to start a conversation) on a 7-point scale, with 1 referring to *very low ability,* 4 referring to *normal ability,* and 7 referring to *very high ability.* Hadley and Rice reported a high correlation between parent and SLP ratings of children's language, although parent ratings were somewhat higher.

All six of the children in the LI3 group were rated below normal on at least four of the 14 items. Comparison to norm-referenced expectations was accomplished through administration of a standardized test, the Structured Photographic Expressive Language Test—Preschool Second Edition (SPELT–P2; Dawson, Eyer, & Fonkalsrud, 2005). We used a cutoff score of 87 (–.86 *SD*) because this cutoff score yielded a sensitivity of 90.6% and a specificity of 100% (Greenslade, Plante, & Vance, 2009). All six children in the LI3 group scored below this cutoff. Although not used for classification, the children in the LI2 group were also administered the parent rating scale and the SPELT–P2.

The 17 children in the TL group were matched to the children in the LI group based on gender and age. There was no prior concern about language development for any of these children. All of the children in the TL group received a rating of ≥4 (i.e., normal ability) on at least 12 of the 14 items, with no item receiving a rating <3. All of the children in the TL group scored ≥88 on the SPELT–P2, consistent with the –.86 *SD* cutoff suggested by Greenslade et al. (2009). Children who scored >1 *SD* above the mean were excluded so that we would be comparing the LI group to children within the average range rather than children with precocious language. This criterion and the requirements for participant matching meant that some of the children in the TL group from Eisenberg et al. (2012) were excluded from the current study. The resulting TL group included 12 of the children from Eisenberg et al. and five additional children. The SPELT–P2 scores and other demographic data of the TL and LI groups are provided in Table 1. The SPELT–P2 scores of the LI2 and LI3 groups are further reported separately.

The TL and LI groups did not differ in age, $F(1, 32) = 0.10$, $p = .75$, $\eta_p^2 = 0.003$. In contrast, the TL group obtained a significantly higher score than the LI group on the SPELT–P2, $F(1, 32) = 44.11$, $p < .001$, $\eta_p^2 = 0.58$. In addition, within the LI group, the LI2 group scored higher on the SPELT–P2 than the LI3 group, $F(1, 15) = 7.58$, $p = .02$, $\eta_p^2 = 0.34$. Based on Cohen (1992), we interpreted the magnitude of effect size (i.e., η_p^2, partial eta-squared) using the following

Table 1. Mean and standard deviation of background measures of the children by group.

	Age in months		*SPELT–P2*[a]		*RIST*[a]	
	M	SD	M	SD	M	SD
TL (N = 17)	41.65	3.62	107.59	4.61	108.06	13.03
LI—All (N = 17)	41.19	3.29	87.63	11.77	107.07	14.73
LI2 (n = 11)	40.00	2.95	91.50	8.42	107.50	11.59
LI3 (n = 6)	41.67	3.72	78.83	9.04	103.17	18.45

Note. TL = children with typical language, LI2 = children who had been previously diagnosed with language impairment at age 2, LI3 = children who were diagnosed with language impairment at age 3 as part of the current study.

[a]Standard scores on the SPELT–P2 (Structured Photographic Expressive Language Test—Preschool Second Edition; Dawson, Eyer, & Folkansrud, 2005) and the RIST (Reynolds Intellectual Screening Test; Reynolds & Kamphaus, 2003) are presented.

criteria: $0.01 \leq \eta_p^2 < 0.09$, small effect size; $0.09 \leq \eta_p^2 < 0.25$, medium effect size; $0.25 \leq \eta_p^2$, large effect size. Thus, both the difference between the TL and LI groups and the difference between the two LI groups on the SPELT–P2 had a large effect size.

All of the children in our study spoke mainstream English; passed a hearing screening at 25 dB for the frequencies 500, 1000, 2000, and 4000 Hz; and had cognitive ability within the typical range as measured by the Odd-Item-Out task of the Reynolds Intellectual Screening Test (RIST; Reynolds & Kamphaus, 2003). The RIST includes both a verbal task and a nonverbal task. Because we were evaluating children with LI, we used only the nonverbal task. The performance of the TL and LI groups on the RIST, as summarized in Table 1, did not differ significantly, $F(1, 32) = 0.01$, $p = .92$, $\eta_p^2 < 0.001$. Based on parent completion of a questionnaire, there was no history or current concern about psychobehavioral, neurological, or physical development for any of the children. All of the children passed the Articulation subtest of the Fluharty Preschool Speech and Language Screening Test—2 (Fluharty, 2001). Socioeconomic status was based on maternal education, with 94% of the mothers having a college degree and 6% having a high school diploma. No group difference occurred with respect to the number of mothers who attended college, $\chi^2(1, N = 34) = 0.00$, $p = 1.00$. The racial distribution based on self-identification by the parent was 60% Caucasian, 15% African American, 9% Asian, and 17% Hispanic.

Materials

Language samples were elicited by asking each child to talk about pictures. There were 15 pictures, each with at least three characters. Eight of the pictures illustrated a problem (e.g., two boys fighting over a bucket in the sandbox and a woman running toward them); the other seven depicted a scene in which the characters were involved in different actions (e.g., a family of four, with the father making toast, the mother pouring milk, one girl at the table waiting to eat, and another girl still in pajamas). Twelve of the 15 pictures were colored line drawings from children's books; the remaining three pictures were photographs from magazines. The pictures were piloted to ensure their suitability for use with young children.

We chose a picture description task, instead of conversation during play, as the context for language sampling because of limitations to conversational language sampling. Although conversational samples have the advantage of reflecting children's functional language skills, they take considerable time to transcribe and analyze. The context and referents for the child's utterances may not always be clear, which makes transcription difficult. In addition, conversational samples are inherently variable in the topics that are talked about and in the adult utterances that are addressed to the child, which may confound the variables of interest. Given these limitations, more structured elicitation procedures about a known context (e.g., picture descriptions) have been suggested to standardize the elicitation procedure across children and make transcription easier and quicker (Dollaghan, Campbell, & Tomlin, 1990; Evans & Craig, 1992). Furthermore, although Lee (1974) used three contexts for DSS—playing with toys, followed by talking about pictures, and then telling a familiar story—these were administered in a fixed order, and most of the samples for her normative data included utterances during picture description and storytelling rather than utterances produced during play. Using the picture description task made our sampling contexts more compatible with those in the DSS analysis and thus allowed us to compare the current results with the findings in Lee.

Procedure

Each child was tested individually by an examiner. The examiners included the second author as well as student research assistants (RAs). Pictures were presented one at a time. To avoid order effects, the pictures were randomized for each child. Each trial began with presentation of the picture. Four elicitation questions, adapted from Leonard, Bolders, and Miller (1976), were asked about each picture. The questions were designed to create uniform opportunities for each child to produce declarative utterances under conditions that obligated full sentences with a subject and predicate (Washington, Craig, & Kushmaul, 1998). The use of elicitation questions has also been shown to yield more varied and advanced language forms than a conversational sample would (Evans & Craig, 1992).

The first, second, and last elicitation questions were kept consistent across all of the pictures. The third elicitation

question, involving a story starter, differed among the pictures. If the child did not respond to one of the elicitation questions, responded with "I don't know," or produced an off-topic utterance, a prompt was given. A list of the elicitation questions and prompts are provided in Appendix A. All responses were audio-recorded for transcription and coding.

Transcription

The samples were transcribed by trained RAs according to Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2010) conventions. To be consistent with the SALT reference database, utterances were segmented into communication units (C-units). C-unit segmentation was also chosen because some of the children produced longer responses involving several independent clauses.

As described by Loban (1963), the C-unit is a subdivision of a larger phonological unit, determined by intonation and pausing, and includes no more than one independent clause as well as any dependent phrasal and clausal constituents. Clauses connected by a coordinating conjunction, such as *and, then, so, but,* and *or,* were divided into separate C-units when the conjoined clause included a subject. Conjoined clauses without a subject were not segmented into a separate C-unit. Embedded clauses, including clauses with a subordinate conjunction, also were not segmented into a separate C-unit. However, clauses in which a subordinating conjunction was used as if it were a coordinating conjunction were segmented into a separate C-unit (e.g., *the girl started to do something because the girl want to dance because there's a wagon*).

Following the conventions of DSS (Lee, 1974), all utterance-initial uses of *and* after C-unit segmentation were eliminated. Furthermore, during development of the PGU measure, we found low reliability for judging the turn-initial conjunctions that were produced after the elicitation question or prompt. For this reason, we decided to eliminate all conjunctions that were produced at the start of turns immediately following the elicitation question or prompt and that did not really conjoin constituents together. For instance, when the examiner asked "What else is happening in this picture," the child might answer "Because they're not sharing." In this case, the utterance-initial *because* was eliminated. It should be noted that C-units can also be utterances without a main clause (i.e., without a subject and/or main verb) when such utterances are preceded and followed by a terminal silence (Hughes, McGillivray, & Schmidek, 1997).

Only complete, intelligible, and on-topic C-units (i.e., only language focused on the pictures in contrast to external comments to the examiner) were included in the analysis. We excluded the questions that children asked about the pictures during the task for several reasons. Questions involve movement or addition of auxiliaries, and young children frequently produce errors on these elements (Rowland, Pine, Lieven, & Theakston, 2005). Because only a few of the children asked questions, including questions in the analysis, could potentially have lowered the grammaticality scores for the few children who did produce questions. In order to avoid this variability, and because the procedure was designed to elicit declarative utterances, we included only the declarative sentences that were produced by each child in the analysis. In addition, replies to examiner requests for repetition or clarification were excluded if the reply was an utterance with ellipsis of the clausal subject and/or verb or if the reply involved an expansion or correction of the child's immediately prior utterance.

Error Coding and Computation

Coding. The errors that children produced in the task were coded as one of the following types. Sample transcripts with examples of each error type are provided in Appendix B.

- *Fragments* were defined as utterances that lacked a verb. These were judged as ungrammatical because the elicitation questions and prompts obligated a complete sentence with a subject and verb. However, utterances with an omitted copula were coded as tense marker errors and were not coded as fragments (see below).
- *Argument structure errors* were defined as omissions of obligatory constituents before or after a verb. Decisions about required postverbal arguments were based on the *Longman Dictionary of Contemporary Usage* (1999–2009). Utterances involving transitive verbs in which the patient occurred in subject rather than in postverbal position (e.g., *the milk spilled*) were not judged as having an argument error. Any omission that could be considered a pragmatically allowable elision was not coded as an argument error.
- *Pronominal form errors* were defined as substitution errors for subject, object, reflexive, and possessive pronouns and possessive determiners. Gender errors were judged based on inconsistencies in the child's utterances and were not based on whether pronoun use agreed with the pictured character. An example of a gender error would be a child who referred to a pictured character as *the dad* and subsequently referred to that same character as *she*.
- *Tense markers errors* were defined as omissions and usage errors for copula, auxiliaries, auxiliary *do,* bound tense markers, and irregular past and third person verb forms. Verbs produced without an inflection, modal, or auxiliary were transcribed as bare verbs, and inappropriate uses of bare verbs were coded as tense marker errors, regardless of the nature of the omission. Uninflected verbs with plural nouns and pronouns (e.g., *They pull the dog*) were not coded as errors unless it was clear from the context that the context obligated a tense marker. Errors were only coded as auxiliary omissions when the child had produced a present or

past participle form that created an obligatory context for the auxiliary. Colloquial uses of *got* (e.g., *He got dirt on his cheek*) were not coded as errors

- *Grammatical morpheme errors* were defined as omissions or substitutions of (a) bound or free nominal morphemes other than pronouns (e.g., plural – *s,* articles), (b) aspect markers (e.g., present participle – *ing*), and (c) prepositions. Errors were only coded as omissions of aspect markers when the child had produced an auxiliary that created an obligatory context for the aspect marker (e.g., *He's drive*ing the car*).
- *Other errors* were defined as any other syntactic error or semantic irregularity that could not be assigned to another error category. Included in this category were lexical errors on content words (i.e., nouns, verbs, and adjectives). As for pronouns, gender substitutions (e.g., *girl* for *boy*) were based on inconsistencies in the child's utterances rather than on whether the word matched the pictured character.

Descriptive measures. We computed PGU, percentage sentence point (PSP), and percentage verb tense usage (PVT) for each child. To compute PGU and PSP, C-units containing one or more error codes were marked as ungrammatical. PGU was calculated by subtracting the number of ungrammatical C-units from the total number of C-units and then dividing by the total number of C-units. The computation of PSP was similar to PGU, except that PSP excluded the C-units without subjects (e.g., *sitting there*) or main verbs (e.g., *he sad*) whereas PGU included these C-units. We computed PSP in order to compare the accuracy of PGU and the DSS sentence point proposed by Lee (1974) in differentiating children with and without LI. Converting the sentence point score into a percentage score allowed us to make the comparison directly. Consistent with the scoring procedures for the DSS sentence point (Lee, 1974), the PGU and PSP measures used in the current investigation coded errors in grammatical morphology, syntax, and semantics. Pragmatic context was considered in determining utterance inclusion, but utterances were not judged for conformity with pragmatic and discourse rules.

To compute PVT, all verb contexts that obligated tense marking were marked. This included contexts for copula, auxiliary *be,* auxiliary *do,* regular past, regular third person singular, irregular past, and third person verb forms. PVT was calculated by subtracting the total number of tense marker errors from the total number of obligatory contexts for tense marking in all C-units and then dividing by the total number of obligatory contexts for tense marking.

Indices of diagnostic accuracy. To compare the diagnostic accuracy of PGU, PSP, and PVT, we computed the sensitivity, specificity, and likelihood ratios for these measures. Sensitivity was computed as the percentage of children with LI who were correctly identified as having LI by PGU, PSP, or PVT. Specificity was computed as the percentage of children with TL who were correctly identified as having TL by PGU, PSP, or PVT. Sensitivity and specificity >80% were considered acceptable, and those >90% were considered good (Plante & Vance, 1994).

Likelihood ratios were calculated from the sensitivity and specificity levels. The positive likelihood ratio (LR+) was calculated as the ratio of true LI to false LI (i.e., sensitivity/[1 – specificity]). A higher LR+ value for a positive test result indicates a higher likelihood that the positive result comes from a child with LI than from a child with TL (Heilmann, Weismer, Evans, & Hollar, 2005). For instance, a PGU of 40% with an LR+ value of 10 would mean that children producing a PGU of 40% are >10 times as likely to have LI than TL. In contrast, the negative likelihood ratio (LR–) is calculated as the ratio of false TL to true TL (i.e., [1 – sensitivity]/specificity). A lower LR– value for a negative test indicates a lower likelihood that the negative result comes from a child with LI than from a child with TL. For instance, a PGU of 60% with an LR– value of 0.02 would mean that children producing a PGU of 60% are <0.02 times as likely to have LI than TL. The value of likelihood ratios can range from 0 to infinity. Dollaghan (2007) suggested that a test should ideally have an LR+ of ≥10 and an LR– of ≤0.10.

To compute sensitivity, specificity, and likelihood ratios, cutoff scores for a positive result were first determined by using the receiver operating characteristic (ROC) curve (C. Brown & Davis, 2006; Sackett, 1991) via SigmaPlot software (Systat Software, 2011). The ROC curve plots the pairs of specificity rates (i.e., hit rate; proportion of LI children who were correctly identified as LI) against the *1 – specificity* rates (i.e., false alarm rates; proportion of TL children who were incorrectly identified as LI) as the cutoff score changes. Thus, the ROC curve analysis automatically calculates pairs of sensitivity and specificity rates for a range of cutoff scores. Following Poll, Betz, and Miller (2010), we chose the score that maximized the diagnostic accuracy, where sensitivity plus specificity divided by 2 is largest, as the cutoff. This procedure avoided the arbitrariness that occurs by setting the cutoff with prescriptive standards (e.g., –1.25 *SD*). Using this cutoff score, we then computed the value of sensitivity, specificity, and likelihood ratios for PGU, PSP, and PVT.

Reliability

We used a consensus procedure (adapted from Shriberg, Kwiatkowski, & Hoffman, 1984) to check the reliability of the transcriptions. Each sample was transcribed by one RA, who was instructed to listen to each utterance a maximum of three times. Utterances that could not be fully transcribed after three listenings were marked as unintelligible and were excluded from the analysis. A second RA then listened to the recorded sample while reading the initial transcription to check the transcription. Transcription for the entire sample was then rechecked by the first author. Discrepancies were discussed and agreement was obtained

on all transcripts. Utterances that could not be resolved were excluded from the analysis.

The same consensus procedure was followed for utterance segmentation, utterance inclusion, and coding. There were no disagreements for utterance segmentation or utterance inclusion. All instances of error coding that could not be resolved were considered to be acceptable and were not coded as errors.

RESULTS

Descriptive Measures of the Language Samples

The descriptive measures of the language samples in the TL and LI groups are presented in Table 2. The data from children with LI who were identified at age 2 (LI2) and those who were identified at age 3 (LI3) are reported separately.

The LI group produced fewer C-units as compared to the TL group, regardless of whether or not utterances without a subject and/or main verb were included in the C-unit count; Fs > 4.19, ps < 0.05, $\eta_p^2 > 0.12$. In addition, the LI group produced shorter C-units in morphemes than the TL group, $F(1, 32) = 6.56$, $p = .02$, $\eta_p^2 = 0.17$. However, the two groups did not differ in their production of C-units in words, $F(1, 32) = 3.15$, $p = .09$, $\eta_p^2 = 0.09$, which supports the finding that calculating utterance length at the level of words may not be sensitive to language ability differences (R. Brown, 1973).

Within the LI group, the LI3 group produced more C-units as compared to the LI2 group when fragments were included, $F(1, 15) = 5.16$, $p = .04$, $\eta_p^2 = 0.26$; but not when fragments were excluded, $F(1, 15) = 1.97$, $p = .25$, $\eta_p^2 = 0.09$. This suggests that the LI3 group produced more fragments than the LI2 group. However, there were no significant differences between the LI2 and LI3 groups in mean length of C-units, Fs < 0.24, ps > 0.88, $\eta_p^2 < 0.002$.

Table 2. Mean and standard deviation of background measures of the picture description task.

	Number of C-units: All		*Number of C-units: DSS*		*$MLCU_w$*[a]		*$MLCU_m$*[a]	
	M	SD	M	SD	M	SD	M	SD
TL	72.41	17.79	69.59	16.59	4.89	0.82	6.06	0.80
LI—ALL	62.00	15.02	52.56	17.10	4.34	1.03	5.19	1.18
LI2	48.50	9.59	43.00	6.86	4.26	0.64	5.15	0.69
LI3	71.00	19.42	58.17	27.46	4.32	1.52	5.14	1.77

Note. Number of C-units: DSS = C-units that were included based on the Developmental Sentence Scoring (Lee, 1974) convention; $MLCU_w$ = mean length of C-units in words; $MLCU_m$ = mean length of C-units in morphemes.

[a]$MLCU_w$ and $MLCU_m$ were computed from all C-units.

Table 3 shows the PGU, PSP, and PVT values by group. Because the participants were composed of boys and girls and ranged in age from 36 to 47 months, we tested whether gender and age affected the performance of PGU, PSP, and PVT before we examined group differences for these measures. One-way analyses of variance (ANOVAs) indicated that the performance of PGU, PSP, and PVT did not vary with gender, Fs < 0.87, ps > 0.71, $\eta_p^2 < 0.004$. Linear regressions showed that age did not significantly account for the variation of PGU, PSP, or PVT; R^2s < 0.014, ps > 0.50, $f^2 < 0.01$. Thus, we did not consider the factors of gender and age in the following analyses.

Next, we compared the differences in PGU, PSP, and PVT between the LI and TL groups. One-way ANOVAs indicated that, as compared to the TL group, the LI group produced lower PGU, $F(1, 32) = 65.69$, $p < .001$, $\eta_p^2 = 0.67$; lower PSP, $F(1, 32) = 52.54$, $p < .001$, $\eta_p^2 = 0.62$; and lower PVT, $F(1, 32) = 29.56$, $p < .001$, $\eta_p^2 = 0.48$. In addition, both the TL and LI groups produced higher PSP than PGU: $F(1, 16) = 15.86$, $p = .001$, $\eta_p^2 = 0.50$ for the TL group; $F(1, 16) = 21.36$, $p < .001$, $\eta_p^2 = 0.57$ for the LI group.

Within the LI group, the LI2 group produced higher PGU than the LI3 group, $F(1, 15) = 7.78$, $p = .01$, $\eta_p^2 = 0.34$. However, the LI2 and LI3 groups did not differ significantly in PSP, $F(1, 15) = 4.12$, $p = .61$, $\eta_p^2 = 0.02$; or in PVT, $F(1, 15) = 2.45$, $p = .24$, $\eta_p^2 = 0.14$.

Indices of Diagnostic Accuracy

We conducted ROC curve analyses to determine the cutoff scores for PGU, PSP, and PVT that would best differentiate between the LI and TL groups. The best cutoff for this was 58.32% for PGU, 67.46% for PSP, and 85.48% for PVT.

Table 3. Mean, standard deviation, and range for percentage grammatical utterances (PGU), percentage sentence point (PSP), and percentage verb tense usage (PVT) by group.

	Mean	SD	*Range*
TL			
PGU	72%	12%	46%–89%
PSP	75%	11%	48%–89%
PVT	91%	8%	71%–100%
LI—All			
PGU	38%	12%	39%–57%
PSP	45%	13%	48%–67%
PVT	66%	16%	32%–85%
LI2			
PGU	43%	9%	30%–57%
PSP	48%	10%	33%–67%
PVT	74%	14%	40%–85%
LI3			
PGU	30%	11%	17%–44%
PSP	38%	13%	22%–50%
PVT	59%	17%	32%–83%

Table 4 presents the indices of diagnostic accuracy of PGU, PSP, and PVT with the cutoff value. PGU, PSP, and PVT all showed good sensitivity to LI, correctly identifying children with LI without missing any (i.e., sensitivity = 100%). However, PGU, PSP, and PVT all demonstrated only fair levels of specificity, meaning that some of the children with TL were misclassified as LI. PGU had a slightly higher specificity level (i.e., 88%) than either PSP or PVT (i.e., 82%). In addition, PGU had a higher LR+ value than either PSP or PVT (8.50 vs. 5.67), although all three measures had an LR+ value <10, which is lower than the level that Dollaghan (2007) suggests.

DISCUSSION

Our first aim was to investigate the diagnostic accuracy of a measure of grammatical accuracy, PGU, for differentiating between children with and without LI. Our other aims were to compare PGU to two less comprehensive measures of grammaticality—PSP, which excludes utterances without a subject or main verb, and PVT, which considers only errors on verb tense marking. We also checked for agreement with the previous study by Eisenberg et al. (2012) because the current group of TL children included 12 of the children from that earlier report.

The 72% mean PGU for the TL group in our study was not statistically different from the 71% PGU previously reported by Eisenberg et al. (2012). At a cutoff score of 58%, PGU showed good sensitivity, >90% level suggested by Plante and Vance (1994), for diagnosing LI. The 88% specificity level at this cutoff was just below the 90% level suggested by Plante and Vance (1994) for diagnosing LI and was above the 80% level suggested by those authors as appropriate for screening purposes.

The 75% mean PSP for the TL group was slightly higher than the 70% of sentences receiving a sentence point reported by Lee (1974) and the 71% level previously reported by Eisenberg et al. (2012) for 3-year-old TL children. Eisenberg et al. (2012) reported comparable results between PGU and PSP (termed PGU-X in Eisenberg et al., 2012). However, in our study, PGU was significantly lower than PSP for the TL children (72% vs. 75%) as well as for the children with LI (38% vs. 45%), and the effect size of this difference was large. The cutoff criterion was similarly lower for PGU than for PSP (58% vs. 67%). This difference was due to the fact that PGU captured a wider range of ungrammatical productions that were excluded from the PSP calculation. We compared the diagnostic accuracy of PGU and PSP. Sensitivity was the same for both measures (100%), >90% level suggested by Plante and Vance (1994), for diagnosing LI. Specificity for PSP was slightly lower than for PGU (82% vs. 88%). This was due to one TL child who fell below the cutoff for PSP but not for PGU.

The sensitivity and specificity of PVT for picture description by 3-year-olds was virtually identical to the diagnostic accuracy previously reported by Bedore and Leonard (1998) for conversational samples from their older group of children: error-free for identifying children with LI, and fair for identifying TL children. There was no difference in sensitivity between PVT and PGU (100% for both). Specificity for PVT was slightly lower than for PGU (82% vs. 88%). This was due to a different child with TL who fell below the cutoff for PVT but not for PGU. PGU was thus comparable in diagnostic accuracy to both PSP and PVT.

The analysis for PGU can also be used to provide qualitative information that can guide goal selection. Here, PGU has an advantage over PSP. Because the PSP analysis, following the DSS conventions, eliminates utterances without a subject and/or main verb, an error analysis of the included utterances would miss certain error types that would be considered in PGU, specifically, subject argument omissions, inappropriate use of fragments, and copula omissions.

PGU has a similar advantage over PVT in considering error types in addition to tense marking errors. Another advantage to PGU over PVT is that it is relatively easier to make judgments about utterance grammaticality than it is to calculate tense marker usage. This makes PGU attractive as a screening measure to identify children who may be in need of further evaluation. In addition, it is not at all clear how to interpret PVT scores below the 85.48% cutoff. Although this was the PVT score that provided the best diagnostic accuracy, this usage rate seems overly high to be clinically relevant given the variability in usage of verb tense markers by TL children at this age (Balason & Dollaghan, 2002; Lahey, Liebergott, Chesnick, Menyuk, & Adams, 1992). Lahey (1994) questioned the applicability of the 90% mastery criteria suggested by R. Brown (1973) to shorter clinical samples and suggested an achievement criteria of 80% usage (Lahey, 1988).

Table 4. Indices of diagnostic accuracy for PGU, PSP, and PVT based on empirically determined cutoff scores of 58.32% for PGU, 67.46% for PSP, and 85.48% for PVT.

	Sensitivity	*Specificity*	*LR+*	*LR–*
PGU	100%	88%	8.50	0.00
PSP	100%	82%	5.67	0.00
PVT	100%	82%	5.67	0.00

Note. LR+ = positive likelihood ratio, LR– = negative likelihood ratio.

Classifying Children as Having an LI

Children in the LI group entered the study in two ways. One group of children (the LI2 group) had been previously diagnosed with an LI and were referred by an SLP. A second

group of children (the LI3 group) were recruited through preschools and pediatricians. The two groups differed in their scores on the standardized language test (i.e., SPELT–P2), with many of the children in the LI2 group scoring within the normal range. In her discussion about using spontaneous speech data for investigating child language, Stromswald (1996) cautioned against setting criteria for utterance exclusion that might bias the results in favor of the predicted hypothesis. Similarly, we felt that eliminating children in the LI2 group who scored above the test cutoff would have biased the results in favor of our prediction that children with LI would perform significantly lower than TL children on PGU.

The reason for the high test performance by the LI2 group is not clear. It could be that the LI2 group of children had a milder degree of LI than the LI3 group at the time of the testing, perhaps because they had been receiving therapy for their language deficits. This possibility is supported by the lower performance of the LI3 group on all three of the language sample measures. It could also be that the SPELT–P2 is not valid for diagnosing children as having an LI if they have previously received language therapy. Whatever the reason, we feel confident that all of the children did, in fact, have an LI for several reasons. The children in both groups were given low ratings about their language by their parents, in keeping with the normativist criteria suggested by Paul and Norbury (2012). The children in both groups were independently diagnosed by an SLP as having an LI and were enrolled in therapy for language, either at the time of the study or immediately afterward. In spite of the difference in test performance, both groups of children scored significantly lower than the TL children on all of the language sample measures, and all of the children classified as LI scored below the cutoff.

Clinical Implications

Measures of diagnostic accuracy are affected by the cutoff score. Many textbooks on language disorders recommend a score at the 10th percentile, or –1.25 *SD*s below the mean, as the clinical cutoff for concluding that a child does not meet age expectations (see, for instance, Fey, 1986; Owens, 2004; Paul & Norbury, 2012). This was also the cutoff used by Tomblin et al. (1996) as the criteria for the Epidemiology of Specific Language Impairment diagnostic system and suggested by Lee (1974) for DSS. This cutoff is, however, arbitrary. Plante and Vance (1994) showed that an empirically determined cutoff can improve diagnostic accuracy. The empirically determined cutoff score for PGU was at a higher score than the –1.25 *SD* cutoff point. Applying the empirically determined cutoff score increased sensitivity without reducing specificity.

The resultant diagnostic accuracy makes PGU an appropriate measure to screen for LI. Plante and Vance (1994) noted that high sensitivity is more important for a screening measure than high specificity because the consequences of missing children who might have LI is a more serious one than overreferring TL children for further evaluation. However, it is also important not to have too high a false positive rate as this imposes a time and economic burden as well as a potential psychological burden for parents.

Quantitative measures of language are valuable for making diagnostic decisions but do not help with selecting treatment goals (Hughes, Fey, & Long, 1992). Although judgments about grammaticality can be made without first coding error type, doing so provides qualitative information that can guide goal selection.

Limitations

The current study used a structured picture sampling procedure with prompts to obligate complete sentences with a subject and predicate. Therefore, the results are not generalizable to language samples that are elicited during play. The pictures used for the current study were from children's books and magazines. Additional investigations with different pictures are needed to determine whether the results are specific to this set of pictures or are applicable to different pictures.

The participants in the current study were classified as TL or LI based on a standardized test score (SPELT–P2) and a parent rating scale. However, the diagnostic accuracy of these measures has not been confirmed for 3-year-olds. Although previous studies of the SPELT–P2 have shown good specificity for 4-year-olds based on an empirically determined cutoff (Greenslade et al., 2009), there has been no study that has looked at the diagnostic accuracy of the SPELT–P2 for younger children or has empirically established an appropriate cutoff score for this age. Hadley and Rice (1993) reported a high correlation between parent and SLP ratings. However, their parent ratings tended to be higher than the SLP ratings. It might, therefore, be the case that some children in the current study were misclassified based on these measures.

Given these caveats, the current data should be used cautiously for setting criteria to diagnose or rule out LI. Longitudinal follow-up data from participants would be important for validating PGU as a screening measure for LI.

Conclusion

In the process of mastering the elements of language, 3-year-old children produce a variety of error types. Although tense marking errors account for a large proportion of errors at this age, children with LI also produce a large number of other errors such as pronominal errors and argument omissions. A more general measure of grammaticality that considers additional aspects of language might, therefore, be useful in screening for LI at this age.

ACKNOWLEDGMENTS

This project was supported by Award Number R21DC009218 from the National Institute on Deafness and Other Communication Disorders (NIDCD) to the first author. The content is solely the responsibility of the authors and does not necessarily represent the official views of NIDCD or the National Institutes of Health. The authors are grateful to the children who participated in the study as well as to their parents who allowed them to participate and to the research assistants who collected and transcribed the samples. Portions of this study were presented at the 2011 Symposium for Research on Child Language Disorders in Madison, WI, and at the 2010 convention of the American Speech-Language-Hearing Association in Philadelphia, PA.

REFERENCES

Balason, D. V., & Dollaghan, C. A. (2002). Grammatical morpheme production in 4-year-old children. *Journal of Speech, Language, and Hearing Research, 45,* 961–969.

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

Brown, C., & Davis, H. (2006). Receiver operating characteristics curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems, 80,* 24–38.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Cohen, J. (1992). Statistics: A power primer. *Psychology Bulletin, 112,* 155–159.

Dawson, J., Eyer, J. A., & Fonkalsrud, J. (2005). *Structured Photographic Expressive Language Test—Preschool, Second Edition*. DeKalb, IL: Janelle Publications.

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

Dollaghan, C. A., Campbell, T. F., & Tomlin, R. (1990). Video narration as a language sampling context. *Journal of Speech and Hearing Disorders, 55,* 582–590.

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

Eisenberg, S., Fersko, T., & Lundgren, C. (2001). Use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology, 10,* 323–342.

Eisenberg, S. L., Guo, L., & Germezi, M. (2012). How grammatical are three-year-olds? *Language, Speech, and Hearing Services in Schools, 43,* 36–52.

Evans, J. L., & Craig, H. K. (1992). Language sample collection and analysis: Interview compared to freeplay assessment contexts. *Journal of Speech and Hearing Research, 35,* 343–353.

Fey, M. E. (1986). *Language intervention with young children*. Needham Heights, MA: Allyn & Bacon.

Fluharty, N. B. (2001). *Fluharty Preschool Speech and Language Screening Test, Second Edition*. Austin, TX: Pro-Ed.

Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implications for assessment and intervention. *American Journal of Speech-Language Pathology, 9,* 151–161.

Greenslade, K. J., Plante, E., & Vance, R. (2009). The diagnostic accuracy and construct validity of the Structured Photographic Expressive Language Test—Preschool: Second edition. *Language, Speech, and Hearing Services in Schools, 40,* 150–160.

Grela, B., & Leonard, L. (1997). The use of subject arguments by children with specific language impairment. *Clinical Linguistics and Phonetics, 11*(6), 443–453.

Hadley, P. A., & Rice, M. L. (1993). Parental judgments of preschooler's speech and language development: A resource for assessment and IEP planning. *Seminars in Speech and Language, 14,* 278–288.

Heilmann, J., Weismer, S. E., Evans, J., & Hollar, C. (2005). Utility of the MacArthur-Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology, 14,* 40–51.

Huang, R., Hopkins, J., & Nippold, M. A. (1997). Satisfaction with standardized language testing: A survey of speech-language pathologists. *Language, Speech, and Hearing Services in Schools, 28,* 12–29.

Hughes, D. L., Fey, M. E., & Long, S. H. (1992). Developmental sentence scoring: Still useful after all these years. *Topics in Language Disorders, 12,* 1–12.

Hughes, D., McGillivray, L., & Schmidek, M. (1997). *Guide to narrative language: Procedures for assessment*. Eau Claire, WI: Thinking Publications.

Hux, K., Morris-Friehe, M., & Sanger, D. D. (1993). Language sampling practices: A survey of nine states. *Language, Speech, and Hearing Services in Schools, 24,* 84–91.

Klee, T., Schaffer, M., May, S., Membrino, I., & Mougey, K. (1989). A comparison of the age–MLU relation in normal and specifically language-impaired preschool children. *Journal of Speech and Hearing Disorders, 54,* 226–233.

Lahey, M. (1988). *Language disorders and language development*. New York, NY: MacMillan.

Lahey, M. (1994). Grammatical morpheme acquisition: Do norms exist? *Journal of Speech and Hearing Research, 39,* 1192–1194.

Lahey, M., Liebergott, J., Chesnick, M., Menyuk, P., & Adams, J. (1992). Variability in children's use of grammatical morphemes. *Applied Psycholinguistics, 13,* 373–398.

Lee, L. L. (1974). *Developmental sentence analysis*. Evanston, IL: Northwestern University Press.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Leonard, L. B., Bolders, J. G., & Miller, J. A. (1976). An examination of the semantic relations reflected in the language usage of normal and language-disordered children. *Journal of Speech and Hearing Research, 19,* 371–392.

Leonard, L. B., Deevy, P., Kutz, R., Chorev, L. K., Owen, A., Polite, E., . . . Finneran, D. (2007). Lexical aspect and use of verb morphology by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 50,* 759–777.

Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. (1997). Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 40,* 741–753.

Loban, W. D. (1963). *The language of elementary school children* (Research Report No. 1). Urbana, IL: National Council of Teachers of English.

Loeb, D. F., Kinsler, K., & Bookbinder, L. (2000, November). *Current language sampling practices in preschools*. Poster presented at the annual convention of the American Speech-Language-Hearing Association, Washington, DC.

Loeb, D., & Leonard, L. (1991). Subject case marking and verb morphology in normally developing and specifically language-impaired children. *Journal of Speech, Language, and Hearing Research, 34,* 340–346.

Longman Dictionary of Contemporary Usage. (1999–2009). Retrieved from http:www.ldoceonline.comabout.html.

McCauley, R. J. (2001). *Assessment of language disorders in children*. Mahwah, NJ: Erlbaum.

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research, 24,* 154–161.

Miller, J., & Iglesias, A. (2010). Systematic Analysis of Language Transcripts (Research version 2010) [Computer software]. Middleton, WI: SALT Software.

Moore, M. E. (2001). Third person pronoun errors by children with and without language impairment. *Journal of Communication Disorders, 34,* 207–228.

Owen, R. E., Jr. (2004). *Language disorders: A functional approach to assessment and intervention* (4th ed.). Boston, MA: Pearson.

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence* (4th ed.). St Louis, MO: Elsevier.

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

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,* 414–429.

Rescorla, L., & Lee, E. (2001). Language impairment in young children. In T. Layton, E. Crais, & L. Watson (Eds.), *Handbook of early language impairment in children: Nature* (pp. 1–55). Albany, NY: Delmar.

Rescorla, L., & Schwartz, E. (1990). Outcome of toddlers with expressive language delay. *Applied Psycholinguistics, 11,* 393–407.

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Screening Test.* Lutz, FL: Psychological Assessment Resources.

Rice, M. L., & Bode, J. V. (1993). GAPS in the verb lexicons of children with specific language impairment. *First Language, 13*(37), 113–131.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairment. *Journal of Speech, Language, and Hearing Research, 53,* 333–349.

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

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

Rowland, C. F., Pine, J. M., Lieven, E. V. M., & Theakston, A. L. (2005). The incidence of error in young children's wh-questions. *Journal of Speech, Language, and Hearing Research, 48,* 384–404.

Sackett, D. (1991). *Clinical epidemiology: A basic science for clinical medicine* (2nd ed.). Boston, MA: Little Brown.

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

Shriberg, L. D., Kwiatkowski, J., & Hoffman, K. (1984). A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research, 27,* 456–465.

Stromswald, K. (1996). Analyzing children's spontaneous speech. In D. McDaniel, C. McKee, & H. S. Cairns (Eds.), *Methods for assessing children's speech* (pp. 23–53). Cambridge, MA: MIT Press.

Systat Sofware, Inc. (2011). SigmaPlot (Version 12.0) [Computer software]. Point Richmond, CA: Author.

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,* 414–429.

Thal, D. J., & Katich, J. (1996). Predicaments in early identification of specific language impairment: Does the early bird always catch the worm? In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (pp. 1–28). Baltimore, MD: Brookes.

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

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

Washington, J. A., Craig, H. K., & Kushmaul, A. J. (1998). Variable use of African American English across two language sampling contexts. *Journal of Speech, Language, and Hearing Research, 41,* 1115–1124.

Watkins, R., & Rice, M. (1991). Verb particle and preposition acquisition in language-impaired preschoolers. *Journal of Speech and Hearing Research, 34*(5), 1130–1141.

APPENDIX A. ELICITATION QUESTIONS AND PROMPTS FOR THE PICTURE TASK

Elicitation Questions

(Prompt if the child did not respond, responded with "I don't know," or produced an off-topic utterance)

1. What is happening in the picture? (PROMPT: POINT TO DIFFERENT PARTS OF PICTURE AND SAY: *Just tell something about the picture*.)
2. What *else* is happening in the picture? (PROMPT: POINT TO DIFFERENT PARTS OF THE PICTURE AND SAY: *Tell me something else about the picture*.)
3. SAY: *Now I'll start the story and you finish it*. PROVIDE STORY STARTER AND SAY: *And then*~ (PROMPT: REPEAT STORY STARTER AND SAY: *And then what happens in the story?*)
4. Tell me one more thing about the story. (PROMPT: POINT OUT PARTS OF THE PICTURE THE CHILD HAS NOT TALKED ABOUT AND SAY: *Just tell me anything else about the picture*.)

Story Starters for Elicitation Question 3

- The boy is trying to get the cookies and then~
- Ohno! The dog ate some of the cake and then~
- The dog is in front of the bus and the bus can't move and then~
- The boys are fighting and here comes the mom and then~
- The children see the cat. The cat is stuck up in the tree and then~
- The daddy is hiding a doll behind his back and then~
- The children are taking the donuts from the bag and then~
- The boy knocked the boxes off the shelf and then~
- The children are trying to wash the dog and then~
- Ohno, the bubbles spilled and then~
- The boys are throwing snowballs and then~
- The dog and the girl have the daddy's shoes and then~
- The little girl is still in her pajamas and then~
- They raked the leaves into a big pile and then~
- The boy is taking his grandma's scissors and then~

APPENDIX B. SAMPLE TRANSCRIPTS WITH CODING FOR THE "DOG EATING CAKE" STORY

Child 1

E What is happening in the picture?
C (um) [EARG] Messing the cake [EU].
E What else is happening in this picture?
C (um) The dog's running on [OTHER] the couch [EU].
E Now I'll start the start the story and you finish it.
E Ohno, the dog ate some of the cake and then~
C [EARG] Is a mess [EU].
E Tell me one more thing about the story.
C (um the) The mommy (was) have [EV] *a [EMOR] broom [EU].
C ((and)) he [EPRO] was mad [EU].

Child 2

E What's happening in the picture?
C (um) The cake [FRAG] [EU].
E What else is happening in this picture?
C (um the) The people are standing.
E Now I'll start the start the story and you finish it.
E Ohno, the dog same of the cake and then~
C (He could) He couldn't get it.
E And tell me one more thing about the story.
C (um) The dog [FRAG] [EU].

Note. E = examiner, C = child, () = maze words, (()) = words and utterances excluded from the analysis. Types of errors: [EARG] = error on argument, [EU] = utterance with one or more errors, [OTHER] = other grammatical or semantic error, [EV] = error on verb tense marking, [EMOR] = error on grammatical morpheme, [EPRO] = error on pronoun, [FRAG] = fragment.

Differentiating Children With and Without Language Impairment Based on Grammaticality

Sarita L. Eisenberg, and Ling-Yu Guo
Lang Speech Hear Serv Sch 2013;44;20-31; originally published online Jul 23, 2012;
DOI: 10.1044/0161-1461(2012/11-0089)