LSHSS

Research Note

Percent Grammatical Responses as a General Outcome Measure: Initial Validity

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

Purpose: This report investigated the validity of using percent grammatical responses (PGR) as a measure for assessing grammaticality. To establish construct validity, we computed the correlation of PGR with another measure of grammar skills and with an unrelated skill area. To establish concurrent validity for PGR, we computed the correlation of PGR with a previously validated measure of grammaticality, percent grammatical utterances (PGU), and examined the extent to which PGR and PGU agreed upon pass/fail decisions for children.
Method: Participants included 79 3-year-olds from mostly middle socioeconomic status homes. Language samples were elicited by asking children to describe 15 pictures in response to 4 questions per picture. To calculate PGU, children's responses to all 4 questions were segmented into communication units, and each communication unit was evaluated for grammatical errors. To calculate PGR, the entire response to just the first question was evaluated for grammatical errors.
Results: PGR scores significantly correlated with a standardized test of grammar (r = .70), but not with a measure of vocabulary (i.e., type–token ratio; r = .11). In addition, PGR scores were significantly correlated with PGU scores (r = .88). Agreement between PGR and PGU was 92% for pass decisions and 94% for fail decisions.
Conclusions: The current study establishes the construct validity of PGR as a measure of grammar and supports the use of PGR as a measure to assess grammaticality.

General outcome measures assess overall proficiency on a broad skill area using standardized procedures that yield a single quantitative score rather than assessing separate components within the skill area (Deno, 2003; Fuchs, 2004; Fuchs & Deno, 1991; McConnell, McEvoy, & Priest, 2002; Phaneuf & Silberglitt, 2003). They are used both to identify low-scoring children so that intervention can be provided and to assess children over time to monitor growth in the skill area and change in response to intervention. Because general outcome measures are meant to be administered frequently, they must be brief and easy to administer.

Fuchs (2004) identified three stages in the development of general outcome measures. Stage 1 investigates technical adequacy—reliability and validity—of the measure itself at a single point in time. Stage 2 compares performance on the measure over several time points to determine whether changes in performance reflect growth in the skill area. Stage 3 investigates the usefulness of the measure for evaluating change in response to an intervention.

The current study is a Stage 1 study of an assessment procedure for measuring grammaticality. Specifically, we addressed several criteria identified by Deno, Mirkin, and Chiang (1982) as important for general outcome measures: (a) that the proposed measure involves an important skill area, (b) that the proposed measure involves a standardized procedure, (c) that the proposed measure is valid for measuring the skill area, and (d) that the proposed measure can substitute for a lengthier, previously validated measure of the skill area.

The Importance of Grammaticality as a General Outcome

The oral language skill Standard 1 of the Common Core Educational Standards for language calls for students to demonstrate command of the conventions of standard English grammar and usage when speaking or writing (National Governors Association Center for Best Practices & Council of State School Officers, 2010). This standard is particularly relevant for children with language impairments (LI), because grammatical errors are a key characteristic of

[a]Department of Communication Sciences and Disorders, Montclair State University, Bloomfield, NJ
[b]Department of Communicative Sciences and Disorders, University at Buffalo, NY
[c]Department of Audiology and Speech-Language Pathology, Asia University, Taichung, Taiwan
Correspondence to Sarita Eisenberg: eisenbergs@montclair.edu

Editor: Shelley Gray
Associate Editor: Monique Mills
Received August 29, 2016
Revision received December 29, 2016
Accepted August 4, 2017
https://doi.org/10.1044/2017_LSHSS-16-0070

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

these children (Leonard, 1998). Measures of grammaticality have been shown to distinguish between preschool children with and without LI (Dunn, Flax, Sliwinski, & Aram, 1996; Eisenberg & Guo, 2013; Souto, Leonard, & Deevy, 2014). Problems with grammaticality persist into the school years. School-age children with LI made more grammatical errors than those with typical development in both spoken and written narrative and expository language samples (Fey, Catts, Proctor-Williams, Tomblin, & Zhang, 2004; Guo & Schneider, 2016; Scott & Windsor, 2000).

Grammatical performance has been identified as an important contributor to later literacy achievement. Children with early LI are at high risk for reading problems (Boudreau & Hedberg, 1999; Justice, Invernizzi, & Meier, 2002). In spite of the importance of early grammatical performance, general outcome measures for preschool oral language have focused on vocabulary and the achievement of word combinations (e.g., Missal, Carta, McConnell, Walker, & Greenwood, 2008; Phaneuf & Silberglitt, 2003). There are no general outcome measures for grammar.

Procedures for Measuring Grammaticality

In children diagnosed with LI, the most frequently noted grammatical deficit is with usage of verb tense markers (Goffman & Leonard, 2000; Rice, Wexler, & Hershberger, 1995). However, children with LI have also been reported to make argument omissions (Grela & Leonard, 1997), pronoun errors (Loeb & Leonard, 1991; Moore, 2001), and omissions and errors on other grammatical morphemes (Leonard, Eyer, Bedore, & Grela, 1997; Polite, Leonard, & Roberts, 2011; Watkins & Rice, 1991). Eisenberg and colleagues (Eisenberg & Guo, 2013; Eisenberg, Guo, & Germezi, 2012), therefore, developed a broad measure of grammaticality, percent grammatical utterances (PGU), that considers a range of grammatical error types.

An important characteristic of general outcome measures, as noted above, is that they use standardized procedures (Fuchs & Deno, 1991; McConnell et al., 2002). Conversational samples are inherently variable in the topics that are talked about and in the adult utterances addressed to the child. A structured picture description task with standardized elicitation was, therefore, chosen instead of conversation during play as the context for calculating PGU (Eisenberg et al., 2012). In this task, children were shown 15 pictures and were asked to respond to four elicitation questions about each picture. Samples were transcribed and segmented into communication units (C-units), utterance units that include no more than one independent clause (Loban, 1963). Each C-unit was evaluated for grammatical errors. Overall, PGU was significantly correlated ($r = .53$) with the Structured Photographic Expressive Language Test–Preschool Second Edition (SPELT-P2; Dawson, Eyer, & Fonkalsrud, 2005), a standardized test of morphology and syntax with high diagnostic accuracy (Eisenberg & Guo, 2011). A PGU cutoff score of 58.32% yielded sensitivity at a 100% level and specificity at an 88% level for differentiating between children with LI and children with typical language development (Eisenberg & Guo, 2013). Similar findings were also observed in school-age children (Guo & Schneider, 2016).

These data support the use of PGU to measure grammaticality. However, administration of the entire picture description task ranged from approximately 1 to 3 min per picture ($M = 1.8$ min), for a total time of 14–40 min ($M = 27$ min, $SD = 7$ min). The task was thus overly long for use as a general outcome measure. To create a measure of grammaticality that would be feasible as a general outcome measure, we shortened the PGU procedure in two ways: (a) To shorten administration time, we used only a single response to each picture for scoring, and (b) to shorter analysis time, we rated the entire response rather than separately rating each C-unit within the response. We termed this assessment procedure as percent grammatical responses (PGR) to distinguish it from the longer PGU measure.

Establishing the Validity of PGR for Measuring Grammaticality

The current study investigated the validity of PGR for assessing grammatical usage. We first evaluated the construct validity of PGR. A construct validity of a measure can be established by examining the extent to which this measure correlates with other measures designed to assess the same skill areas (i.e., convergent validity) and the extent to which this measure does not correlate with other measures designed to assess different skill areas (i.e., divergent validity; Hinkin, 1998). To this end, we computed the correlation of PGR with a measure of grammar, the SPELT-P2 (Dawson et al., 2005), and with a measure of vocabulary, the type–token ratio (TTR; Miller & Iglesias, 2010; Templin, 1957). We then investigated the concurrent validity of PGR by comparing the score from this briefer assessment procedure to the score computed for the longer previously validated version of the task (i.e., PGU). To this end, we first evaluated the extent to which the PGR score was correlated with the PGU score. In addition, because general outcome measures are used to identify the need for intervention based on low performance, we also examined agreement between PGR and PGU for making pass/fail decisions, which provided further evidence for the concurrent validity of PGR. If PGR was highly correlated with PGU and was consistent with PGU in making clinical decisions (i.e., pass/fail), then clinicians could use the shorter PGR procedure to assess grammaticality.

We asked the following questions:

1. Is PGR correlated with SPELT-P2, a measure of grammar, and with TTR, a measure of vocabulary?
2. Would the PGR score (based on one response per picture) be significantly correlated with and yield the same pass/fail decisions as the PGU score (based on four responses per picture and segmented into C-units)?

Method

Participants

There were 79 participants (39 girls, 40 boys) ranging in age from 3;0 to 3;11 (years; months, M = 3;5, SD = 0;3). Participants were given a standardized language test, the SPELT-P2 (Dawson et al., 2005), and parents were asked to rate their child's language. The mean standard score of SPELT-P2 across children was 103.84 (SD = 12.64, range 65–125). However, children were included regardless of their performance on the SPELT-P2 (i.e., whether or not they scored within typical range) so that the sample would include a wide range of PGR/PGU scores. All children were mainstream English-speaking. This was both based on parental report developed by the first author and confirmed by checking parent language for use of dialect features (see Washington, Craig, & Kushmaul, 1998).

All children 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 (Reynolds & Kamphaus, 2003). All children passed the Articulation Subtest of the Fluharty Preschool Speech and Language Screening Test–Second Edition (Fluharty, 2001). As part of the protocol, a parent-elicited conversational language sample was collected from each child during free play. All children had at least 78% completely intelligible utterances in the conversational language sample. Socioeconomic status was based on maternal education, with 94% having a college degree or higher and 6% having a high school diploma. Racial/ethnic distribution based on self-identification by the parent was 60% Caucasian, 15% African American, 14% Hispanic, and 11% Asian. The information regarding socioeconomic status and racial/ethnic distribution was obtained from the parent report as well.

Participants were recruited through nursery school programs, pediatricians, and speech-language pathologists in suburban New Jersey as well as through online announcements. Approval for this research was granted by the Montclair State University Institutional Review Board.

Stimuli

The examiner-elicited language samples were collected by asking the child to talk about 15 pictures (i.e., the picture description task). Each of the pictures included at least two people and had at least one other character that was either a person, animal, or doll. Twelve of the 15 pictures were colored line drawings that had been found in out-of-print children's books. The other three pictures were photographs from magazine advertisements. In some of the pictures, the characters were involved in a problem situation. In other pictures, the characters were involved in thematically related actions that did not involve a problem (see Appendix A for a description of each of the 15 pictures).

Procedure

Each child was tested individually by an examiner for the picture description task. The examiners included the second author as well as student research assistants. The order of pictures in the picture description task was randomized by allowing the child to select the next picture to talk about. There were four elicitation questions, adapted from Leonard, Bolders, and Miller (1976), for 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 et al., 1998).

The first, second, and last elicitation questions were the same for all 15 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 or responded with "I don't know," or produced an off-topic utterance, a follow-up prompt was given. A list of the elicitation questions and prompts are provided in Appendix B. All responses were audio-recorded for transcription and coding.

Transcription

The samples from the picture description task were transcribed by trained research assistants according to Systematic Analysis of Language Transcripts conventions (Miller & Iglesias, 2010). Responses were segmented into C-units, defined by Loban (1963) as including no more than one independent clause together with any dependent phrasal and clausal constituents. 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). Responses to a given elicitation question or follow-up prompt could contain one or more C-units.

Conjoined clauses introduced by a coordinating conjunction (e.g., *and, so, but,* and *or*) were divided into separate C-units only when the conjoined clause included a subject. Conjoined clauses without a subject or embedded clauses introduced with a subordinating conjunction (e.g., *because, after,* and *if*) were not segmented into a separate C-unit.

Error Coding and Analysis

We computed PGU and PGR from the picture description task. In order to establish the divergent validity of PGR, we also computed TTR, a vocabulary measure, from the conversational language sample. Note that the conversational language samples were separate from the picture description task.

Computing PGU

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.

Because the elicitation procedure was designed to elicit declarative utterances and few children produced questions in response to the prompts, only the declarative sentences produced by each child were included in the analysis. 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.

All C-units with one or more errors in grammatical morphology and syntax were marked as ungrammatical. Errors included omissions of obligatory constituents, such as the subject or object, omissions and substitution errors involving grammatical morphemes and pronouns, and other syntactic errors. Production of fragments in response to questions and prompts that obligated a complete sentence with a subject and predicate was also 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.

Computing PGR

PGR was calculated differently. For this calculation, only the response to the first elicitation question (i.e., *What is happening in this picture?*) was rated for grammaticality. In addition, only one rating of grammaticality was assigned for the entire response to the first elicitation question, regardless of the number of C-units. PGR was then calculated by adding up the number of responses that were completely grammatical and dividing this by the total number of scorable responses (15 for all of the children). A comparison between PGU and PGR scoring is provided in Appendix C.

Computing TTR

TTR has been considered a measure of lexical diversity, reflecting children's vocabulary skills (Miller & Iglesias, 2010; Templin, 1957). We computed TTR based on the first 50 utterances in the conversational language samples for each child. TTR was calculated by dividing number of different words by total number of words in 50 utterances.

Reliability

To check the reliability of transcription of the picture description task, a consensus procedure (adapted from Shriberg, Kwiatkowski, & Hoffman, 1984) was used. Each sample was transcribed by one research assistant, who was instructed to listen to each utterance a maximum of three times. Utterances that could not be fully transcribed after listening three times were marked as unintelligible and excluded from the analysis. A second research assistant 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. Discrepancies 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. Any discrepancies in error coding that could not be resolved were considered to be acceptable and were not coded as errors. The reliability of transcription and utterance segmentation/inclusion for conversational language samples was also checked using the consensus procedure (see Eisenberg & Guo, 2015).

To further check the reliability of coding for PGR and PGU, 18% (14/79) of the transcripts were randomly selected and independently recoded by a trained research assistant for the occurrence of errors. Interrater agreement for these transcripts was 94% for PGR and 97% for PGU.

Statistical Analyses

To establish the construct validity (i.e., convergent and divergent validity) of PGR, we used Pearson product–moment correlations to evaluate the extent to which PGR correlated with the raw scores of the SPELT-P2, a related measure of grammar, and with TTR, an unrelated measure. To establish the concurrent validity of PGR, we first used Pearson product–moment correlation to determine the extent to which PGR correlated with PGU. We then evaluated the degree of agreement for the pass/fail decisions between the PGR and PGU scores using the receiver operating characteristic (ROC) curve (Sacket, Haynes, Tuqwell, & Guyatt, 1991) in the software SigmaPlot 12.0 (Systat Software, Inc., 2011). In the analysis, PGU was set as the "reference standard" (Dollaghan & Horner, 2011). Pass/fail decisions for children were first made using a previously established cutoff for PGU of 58.32% (Eisenberg & Guo, 2013). The ROC curve analysis then automatically computed the degree of agreement on the pass/fail decisions between PGR and PGU using a range of cutoff scores for PGR. Following Sackett (1991), we chose the cutoff score of PGR that maximized the degree of agreement on pass/fail decisions between PGR and PGU.

Results

Preliminary Analysis

The mean scores for PGR and PGU are presented in Table 1. One-way repeated-measures analysis of variance indicated that PGR was significantly lower than PGU, $F(1, 78) = 28.74$, $p < .001$, $\eta_p^2 = .269$. That is, grammaticality based on the entire response to only the first elicitation question (PGR) was lower than grammaticality based on the total C-units produced in response to all four elicitation questions (PGU).

Construct Validity of PGR

After children's age was partialled out, the PGR score was significantly correlated with the raw score of

Table 1. Means (standard deviations) and ranges of PGR, PGU, test scores, and type–token ratio.

Measures	# C-units	MLCU-m	*M* (*SD*)	Range
PGR	—	—	55% (24%)	8%–93%
PGU	69.71 (15.38)	5.98 (1.07)	61% (20%)	7%–92%
SPELT-P2	—	—	21.71 (6.28)	4–33
TTR50	—	—	0.46 (0.05)	0.35–0.60

Note. # C-units = total number of C-units for the full sample; MLCU-m = mean length of C-units in morphemes for the full sample; PGR = percent grammatical responses; PGU = percent grammatical utterances; SPELT-P2 = raw score for Structured Photographic Expressive Language Test–Preschool Second Edition; TTR50 = type–token ratio in 50 utterances from conversational language samples.

SPELT-P2 ($r = .70$, $p < .001$), but not with TTR computed from the conversational language sample ($r = .11$, $p = .39$). This means that PGR was correlated with a measure of grammar, but now with a measure of vocabulary. Although it was not the primary purpose of this study, we further evaluated the correlations between the PGU score and those two measures. After children's age was partialled out, the PGU score was significantly correlated with the raw score of SPELT-P2 ($r = .79$, $p < .001$), but not with type–token ($r = .01$, $p = .95$), a pattern similar to PGR.

Concurrent Validity of PGR

After children's age was partialled out, the PGR score was significantly correlated with the PGU score ($r = .88$, $p < .001$). Next, we evaluated the extent to which PGR and PGU agreed upon pass/fail decisions using the ROC curve analysis. To this end, we first adopted the previously established cutoff for PGU of 58.32% (Eisenberg & Guo, 2013) to classify children. Forty-eight children scored above this cutoff (i.e., pass), and 31 children scored below this cutoff (i.e., fail). The ROC curve analysis further showed that, when a PGR cutoff of 51.67% was used to make pass/fail decisions, the agreement between PGR and PGU was 92% (44/48) on pass decisions and 94% (29/31) on fail decisions.

Discussion

Grammaticality is an important outcome included in the Common Core Educational Standards for language. There is currently no general outcome measure for this skill area. In this report, we proposed PGR as a possible general outcome measure. PGR meets several characteristics that are important for general outcome measures: It involves a brief standardized procedure that yields a quantitative score. In this study, we investigated the validity of PGR for measuring grammaticality.

We established the construct validity of PGR by showing that PGR was correlated with an independent measure of grammar (SPELT-P2), but not with a measure of vocabulary (TTR). This result indicates that PGR is a valid measure of children's grammatical skills. We also investigated concurrent validity of PGR by comparing PGR to a lengthier and more time-consuming assessment procedure that was previously shown to be valid for measuring grammaticality (PGU). PGR scores were highly correlated with PGU scores, although PGR scores were, in general, lower than PGU scores. In addition, there was substantial agreement between PGR and PGU in making pass/fail decisions when we used a cutoff of 51.67% for PGR and 58.23% for PGU. Plante and Vance (1994) argued that different assessments require different cutoffs rather than using a single cutoff across assessments. Given that PGR scores were lower than PGU, it made sense that a lower cutoff would be needed for PGR.

Why was PGR generally lower than PGU? Consider an example in which the child produces three C-units in response to the first question (e.g., Examiner: *What's happening in this picture?* Child: *The boys are mad. They are not sharing. They fighting*). In this example, two of the C-units are grammatical, whereas one is not (i.e., *They fighting*). The PGR score would be 0% (0/1) because any error would make the whole response ungrammatical regardless of the number of C-units in the response. The PGU score in this example, however, would be 67% (1/3), because PGU considers the accuracy of each scorable C-unit. Thus, the difference between PGR and PGU in this study may have resulted from the nature of the scoring procedure. Despite the difference, the high correlation between PGR and PGU suggests that children's relative ranking was quite consistent between PGR and PGU.

Clinical Implications

One purpose for general outcome measures is to determine the need for intervention without having to perform a lengthy assessment. This requires that the general outcome measure yields the same pass/fail decisions as the lengthier assessment. Grammaticality as measured by PGR had sufficient agreement with PGU to be useful for this purpose when a cutoff score of 51.67% was used for PGR. This cutoff resulted in a low rate (6%) of disagreement about failure (i.e., performance below the cutoff) relative to the PGU reference standard while also minimizing the rate (8%) of disagreements about passing (i.e., performance

above the cutoff). The standardized elicitation procedure, the short amount of time needed for collecting the sample, and the ability to rate grammaticality of responses without first segmenting them into C-units suggest that PGR could potentially be usable as a general outcome measure of grammaticality.

Limitations

The current study used responses from a larger data set—responses to the first of four elicitation questions about each of a set of 15 pictures. It needs to be verified whether the results would be the same if the elicitation procedure had involved just the one elicitation question (*What is happening in the picture?*). We also do not know how long it would take to administer the task with just one elicitation question and a single response.

The current study focused on PGR performance at a single point in time. However, general outcome measures are used for repeated measures to track change over time and to determine whether change is occurring in response to intervention. Further studies are needed to determine whether PGR is sensitive to differences over time.

General outcome measures require a standardized procedure. However, the pictures that we used may not be available for all clinicians, and we do not know whether use of different pictures would affect PGR scores. In addition, such repeated measures require alternate forms. Further studies that examine the effect of pictures on PGR and the equivalence of PGR elicited with different picture sets are needed.

Moreover, the range of socioeconomic status for the participants (as determined by maternal education level) was not fully representative. This is important because socioeconomic status has been shown to influence performance on language assessments (Rowe, 2008). The results may not, therefore, be generalizable to children whose mothers have lower educational levels.

Concluding Thoughts

Previous studies have established that PGU is a valid measure for grammaticality. The current study established the construct and concurrent validity for an alternative briefer measure of grammaticality (PGR). PGR is highly correlated with PGU and yields pass/fail decisions that were consistent with the longer, more time-consuming PGU. This suggests that PGR might be useful as a general outcome measure, because it involves a standardized procedure for assessing grammaticality that takes a shorter amount of time for eliciting and analyzing responses. However, general outcome measures are used not only for determining the need for intervention at a single point in time but also for tracking change over time. Further investigation is needed in order to determine whether PGR is sufficiently stable and sensitive to change over time to be used for this latter purpose as well. To allow for repeated administrations, alternative forms of the task, using different pictures, will also need to be developed, and the reliability of these alternative forms will need to be determined.

Acknowledgments

The data for this project were collected with support from Award R21DC009218 from the National Institute on Deafness and Other Communication Disorders to the first author. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health. The authors are grateful to the children who participated, as well as to their parents who allowed them to participate, and to the research assistants who collected and transcribed the samples.

References

Bank Street College of Education. (1968a). *"Give me my pail!"* New York, NY: MacMillan.

Bank Street College of Education. (1968b). *The cat in the tree*. New York, NY: MacMillan.

Boudreau, D. M., & Hedberg, N. L. (1999). A comparison of early literacy skills in children with specific language impairment and their typically-developing peers. *American Journal of Speech-Language Pathology, 8,* 249–260.

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

Deno, S. L. (2003). Developments in curriculum-based measurement. *Journal of Special Education, 37,* 184–192.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36–45.

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

Dunn, 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. L., & Guo, L. (2011, November). *Percent grammatical utterances produced by LI and typical children.* Poster preseted at the American Speech-Language-Hearing Association convention, San Diego, CA.

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

Eisenberg, S. L., & Guo, L. (2015). Sample size for measuring grammaticality in preschool children from picture-elicited language samples. *Language, Speech, and Hearing Services in Schools, 46,* 141–153.

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

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

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

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33,* 188–192.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488–500.

Goffman, I., & 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.

Goodglass, H., Kaplan, E., & Barresi, B. (2000). *Boston Diagnostic Aphasia Examination–Third Edition*. Austin, TX: Pro-ed.

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

Guo, L., & Schneider, P. (2016). Differentiating school-aged children with and without language impairment using tense and grammaticality measures from a narrative task. *Journal of Speech, Language, and Hearing Research, 59,* 317–329.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 2*(1), 104–121.

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

Justice, L. M., Invernizzi, M. A., & Meier, J. D. (2002). Designing and implementing an early literacy screening protocol: Suggestions for the speech-language pathologist. *Language, Speech, and Hearing Services in Schools, 33,* 84–101.

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

McConnell, S. R., McEvoy, M. A., & Priest, J. S. (2002). "Growing" measures for monitoring progress in early childhood education: A research and development process for individual growth and development indicators. *Assessment for Effective Intervention, 27,* 3–14.

Miller, J., & Iglesias, A. (2010). Systematic Analysis of Language Transcripts (SALT), Research Version [Computer Software]. Madison, WI: SALT Software, LLC.

Missal, K. N., Carta, J. J., McConnell, S. R., Walker, D., & Greenwood, C. R. (2008). Using individual growth and development indicators to measure early language and literacy. *Infants and Young Children, 21,* 241–253.

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

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards initiative*. Retrieved from http://www.corestandards.org

Phaneuf, R. L., & Silberglitt, B. (2003). Tracking preschoolers' language and preliteracy development using a general outcome measurement system. *Topics in Early Childhood Special Education, 23,* 114–123.

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

Polite, E. J., Leonard, L. B., & Roberts, F. D. (2011). The use of definite and indefinite articles by children with specific language impairment. *International Journal of Speech-Language Pathology, 13*(4), 291–300.

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

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

Robinson, H. M., Monroe, M., & Artley, A. S. (1962a). *Before we read*. Glenview, IL: Scott Foresman.

Robinson, H. M., Monroe, M., & Artley, A. S. (1962b). *We read pictures*. Glenview, IL: Scott Foresman.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language, 35,* 185–205.

Sacket, D. L., Haynes, R. B., Tuqwell, P., & Guyatt, G. H. (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.

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

Systat Software, Inc. (2011). *SigmaPlot® 12.0*. Point Richmond, CA: Author.

Templin, M. C. (1957). *Certain language skills in children: Their development and interrelationships*. Child Welfare Monograph No. 26. Minneapolis, MN: University of Minnesota Press.

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.

Ziefert, H. (1987). *Jason's bus ride*. New York, NY: Puffin Books.

Appendix A

Pictures

1. BREAKFAST: table with bowls, glasses, spoons, and cereal box; woman pouring juice into a glass; girl holding spoon; another girl in pajamas yawning and rubbing her eyes; man putting bread in toaster
2. BUBBLES: boy leaning over wagon; girl tripping over his feet with bubble jar and wand flying out of her hand; bucket of water spilling over; another girl sitting down holding bubble jar and wand
3. BUS: dog in front of bus; man and woman pulling on dog's leash; driver and boy in front of bus (Ziefert, 1987, pp. 19–20)
4. CAKE: partially eaten cake on table with footprints leading to dog under couch; woman holding a broom and boy crying, both facing the couch; woman with boy holding present and woman with girl holding present entering the room through an open door
5. CAT: cat in tree; boy and girl looking up at the tree (Bank Street College of Education, 1968b, pp. 1–2)
6. COOKIE: boy on stool that is tipping over, reaching for cookie in cookie jar on shelf and holding cookie in other hand; girl reaching for cookie; woman standing by overflowing sink drying a plate (Goodglass, Kaplan, & Barresi, 2000)
7. DOLL: girl running out of house toward man; man kneeling down with one arm reaching toward girl and the other holding a doll behind his back
8. DONUT: woman at wheel of car; donut bag on front seat; boy in front seat handing donut to girl in back seat; both children looking at woman with laughing expressions on their faces
9. DRESSING: man looking in mirror and buttoning his shirt; boy holding a tie; girl with a man's shoe on each hand; another girl with man's hat on her head; dog with slipper in his mouth (Robinson, Monroe, & Artley, 1962b, p. 23)
10. LEAVES: girl putting leaves in bag; another girl and a boy playing in the leaves; man and boy raking leaves; another boy holding a rake and waving at woman in car; dog barking at squirrel in tree
11. SANDBOX: two boys in sandbox tugging at a bucket; another bucket in the sand; woman running over to the sandbox (Bank Street College of Education, 1968a, pp. 3–4)
12. SCISSORS: woman sitting in chair and knitting; boy on his hands and knees taking scissors out of basket at side of chair; girl and man kneeling behind the chair, man with finger to lips
13. SNOW: man and girl building a snowman; another girl shoveling snow; two boys throwing snowballs at each other
14. SUPERMARKET: boy reaching for box on top shelf and several boxes falling down; girl reaching for fruit in bin; another girl with shopping cart looking at her and about to push cart into the boy; woman with list in her hand looking at the shelves of boxes (Robinson, Monroe, & Artley, 1962a, p. 23)
15. WASHING DOG: dog jumping out of tub and splashing a girl; boy with a bucket of water (Robinson, Monroe, & Artley, 1962b, p. 2)

Appendix B

Prompts for the Picture Task (Eisenberg & Guo, 2013)

Elicitation question (alternative 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 Prompt 3:

- BREAKFAST: The little girl is still in her pajamas and then ~
- BUBBLES: Oh no, the bubbles spilled and then ~
- BUS: The dog is in front of the bus and the bus can't move and then ~
- CAKE: Oh no! The dog ate some of the cake and then ~
- CAT: The children see the cat. The cat is stuck up in the tree and then ~
- COOKIE: The boy is trying to get the cookies and then ~
- DOLL: The daddy is hiding a doll behind his back and then ~
- DONUT: The children are taking the donuts from the bag and then ~
- DRESSING: The dog and the girl have the daddy's shoes and then ~
- LEAVES: They raked the leaves into a big pile and then ~
- SANDBOX: The boys are fighting and here comes the mom and then ~
- SCISSORS: The boy is taking his grandma's scissors and then ~
- SNOW: The boys are throwing snowballs and then ~
- SUPERMARKET: The boy knocked the boxes off the shelf and then ~
- WASHING: The children are trying to wash the dog and then ~

Prompts reprinted from Eisenberg and Guo (2013), http://lshss.pubs.asha.org

Appendix C

Sample Scoring of Two Pictures

Elicitation Questions and Child Responses	PGU	PGR
Picture 1: 1 C-unit in response to first elicitation question		
E What's happening in the picture?		
C A dog don't[EW:doesn't] want a bath because he *is all dirty.	0	0
E What else is happening in this picture?		
C A girl *is getting wet.	0	—
E Now I'll start the story and you finish it.		
E The children are trying to wash the dog and then ~		
C A dog run away.	0	—
E Tell me one more thing about the picture.		
C That boy need/*3s more water.	0	—
Picture 2: 2 C-units in response to first elicitation question		
E What's happening in the picture?		
C The boy/s are not happy.	1	0
C They *are mad because they want that red bucket.	0	
E What else is happening in this picture?		
C A mom has to stop them for[EW:from] get/ing mad.	0	—
E Now I'll start the story and you finish it.		
E The boys are fighting and here comes the mom and then ~		
C A mom had to run to them and be friend/s again.	0	—
E Tell me one more thing about the picture.		
C They *are go/ing to fill that toy.	0	—

Note. C-units = communication units; PGU = percent grammatical utterances; PGR = percent grammatical responses; E = examiner; C = child; EW = word error; * = omission; / = morpheme boundary; 3s = 3rd person singular verb morpheme.