JSLHR

Research Note

Tracking the Growth of Tense and Agreement in Children With Specific Language Impairment: Differences Between Measures of Accuracy, Diversity, and Productivity

Laurence B. Leonard,[a] Eileen Haebig,[a,b] Patricia Deevy,[a] and Barbara Brown[a]

Purpose: Composite measures of children's use of tense and agreement morphology differ in their emphasis on accuracy, diversity, or productivity, yet little is known about how these different measures change over time. An understanding of these differences is especially important for the study of children with specific language impairment, given these children's extraordinary difficulty with this aspect of grammar.

Method: We computed 3 types of composite scores from spontaneous speech samples obtained from 17 preschoolers with specific language impairment before, during, and after their participation in a language intervention study. These measures were the Finite Verb Morphology Composite (a measure of accuracy), the Tense Marker Total (a measure of diversity), and the Productivity Score (a measure of productivity).

Results: The 3 measures differed in their growth trajectories. Sample size did not alter the linear or quadratic nature of growth of any composite, although it did affect the absolute values found for the Tense Marker Total and Productivity Score.

Conclusion: Even when sample size is controlled, early growth can be seen in tense and agreement accuracy with relatively limited diversity and productivity, whereas later growth in diversity and productivity can occur with very little change in accuracy.

Approximately 7% of the preschool-age population exhibits a significant deficit in language ability without showing other weaknesses that would lead to a diagnosis such as hearing impairment, intellectual disability, neurological impairment, or autism spectrum disorder (Tomblin et al., 1997). Although falling outside these categories, these children nevertheless constitute a heterogeneous population (Leonard, 2014a). At present, these children are most frequently referred to in the literature as children with specific language impairment (SLI; Bishop, 2014; Leonard, 2014b). However, this term has not enjoyed universal acceptance because it implies that no weaknesses other than language are evident in these children. For many children, this is not the case; nonverbal IQs are often below average (even if safely above the intellectual disability range), and motor skills can also be marginal in some children. Because the field has not yet reached consensus on an alternative clinical label, SLI will be the term used here.

One hallmark of SLI in children acquiring English is an extended period of inconsistency in the use of tense/agreement morphemes. These morphemes include the inflections third-person singular *–s,* past tense *–ed,* and the function words copula *is, are, am, was,* and *were*; auxiliary *is, are, am, was,* and *were*; and auxiliary *do, does,* and *did*. All of these morphemes have been found to be vulnerable in these children. Although they differ slightly in their point of emergence, these morphemes can be characterized as a constellation (Rispoli, Hadley, & Holt, 2012).

Given the common grammatical functions and similar ages of emergence of these tense/agreement morphemes, composites of these morphemes have been created by investigators and used in comparisons between children with SLI and their typical peers. One such composite—constructed in slightly different ways by different investigators (Goffman & Leonard, 2000; Rice, Wexler, & Hershberger, 1998)—is the Finite Verb Morphology Composite (FVMC). This composite is simply a percentage of tense/agreement morpheme use in obligatory contexts with all morphemes and contexts

[a]Purdue University, West Lafayette, IN
[b]Louisiana State University, Baton Rouge, LA
Correspondence to Laurence B. Leonard: xdxl@purdue.edu

Editor: Sean Redmond
Associate Editor: Filip Smolik
Received November 15, 2016
Revision received March 28, 2017
Accepted July 26, 2017
https://doi.org/10.1044/2017_JSLHR-L-16-0427

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

combined. For the preschool years, this measure shows clear developmental growth with sample sizes of 50 utterances or more (Goffman & Leonard, 2000). The FVMC shows lower use by children with SLI than by younger typically developing children matched for mean length of utterance (Leonard, Eyer, Bedore, & Grela, 1997; Rice & Wexler, 1996). Furthermore, by exhibiting good sensitivity and specificity, this measure provides high levels of diagnostic accuracy (Bedore & Leonard, 1998; Rice & Wexler, 2001).

Two other composites are the Tense Marker Total (TMT) and the Productivity Score (PS) developed by Hadley and her colleagues (Hadley & Holt, 2006; Hadley & Short, 2005). These measures are sensitive to the distinction between productions reflecting "direct activation"—that is, word–morpheme combinations that co-occur frequently in the input and might be learned on a rote basis—and productions reflecting "grammatical encoding," which are word–morpheme combinations that are more likely to be constructed productively from grammatical representations (Rispoli et al., 2012). This distinction is operationalized by excluding all instances in which contracted copula or auxiliary forms appear with pronoun subjects (e.g., *it's, that's*). In this respect, both the TMT and the PS differ from the FVMC because the latter includes such potentially memorized forms. Because these forms are frequent, the FVMC might overestimate a child's true ability with tense/agreement morphology.

The TMT is computed by crediting the child with 1 point for the use of each of the 15 tense/agreement morphemes noted earlier. It thus serves as a measure of tense/agreement diversity. Such diversity is not ensured in the FVMC because, as a measure reflecting all tense/agreement morphemes and obligatory contexts combined, the FVMC can be unduly influenced by proficiency in only one or two morphemes for which there is an abundance of contexts.

The PS also reflects diversity, but—true to its name—it emphasizes productivity. For example, a child can earn 5 points by using third-person singular *–s* with five different lexical verbs (e.g., *runs, jumps, likes, wants, plays*) and can earn 5 points by using copula *is* with five different subject nouns (e.g., *girl is, boy is, dog is, monkey is, dolly is*). The computation of the PS (described in greater detail in the Method section) allows for a maximum score of 25. The emphasis on productivity reveals a sharp difference from the FVMC. In the latter, a child's composite will increase with five correct productions of the third-person singular inflection *–s* even if all productions involve the same verb (e.g., *likes*). Only 1 point would be awarded in the case of the PS.

The TMT and PS composites were originally created to track the early development of tense/agreement use in young typically developing children (Hadley & Holt, 2006) and those who might be at risk for later language difficulties (Hadley & Short, 2005). Subsequently, investigators have applied these measures to the speech of 4- and 5-year-old preschoolers with SLI along with their typically developing peers. As with the FVMC, the TMT and PS composites reveal clear vulnerabilities in children with SLI and provide good sensitivity and specificity (Gladfelter & Leonard, 2013).

The FVMC, TMT, and PS composites differ not only in their relative emphasis on accuracy, diversity, and productivity, respectively, but also on their relative dependence on speech sample size. The FVMC takes the number of obligatory contexts into account, and therefore changes in its value are not likely to be dramatic with increases or decreases in the size of the sample. In contrast, both the TMT and the PS are additive measures; a significant increase in sample size is likely to result in an increase in both of these measures. Interpretation of a child's TMT or PS composite then must be done with an eye toward the number of utterances in the speech sample.

In this study, we ask how these three composite measures develop in children with SLI during a period of intervention for one particular tense/agreement morpheme, third-person singular *–s*. Because the tense/agreement morphemes in these composites represent a constellation of forms that develop at approximately the same point in time, increases in these composites were expected during the course of intervention. However, given the differences in the manner in which the FVMC, TMT, and PS are constructed, there is a reason to assume that their patterns of growth will differ, even when the sample size is controlled. Because the FVMC can be heavily influenced by the use of frequently occurring forms that might have been acquired through rote learning and produced via direct activation independent of any true grammatical learning, this composite might show an initial period of sharp growth (as routinized forms are being acquired) followed by a slower period of growth when change reflects primarily the acquisition of morphosyntactic forms produced through grammatical encoding. In contrast, the TMT and PS composites do not include forms that are likely to be the result of direct activation. For this reason, their pattern of growth may be less steep initially and more linear in nature throughout the intervention period.

Without a clear understanding of how growth proceeds with these three composites, misinterpretation of children's gains might easily occur. For example, an early increase in the FVMC might be viewed as a developmental spurt, but much more modest gains might be seen for the same period in the TMT and PS. Similarly, in treatment, a large early increase in the FVMC might be interpreted as a sign of significant early generalization to nontargeted tense/agreement morphemes, whereas a smaller increase in the PS at the same time point might be viewed as a less significant occurrence. Yet, the two may reflect equally valid developmental events, differing only in the way accuracy and productivity change over time.

Along with providing insight into the differences among the three composites, this study can provide us with a better understanding of how growth in these measures might be tied to the age and clinical status of the children. Rispoli et al. (2012) calculated PS composites for spontaneous speech samples obtained from young typically developing children in 3-month intervals from 21 to 30 months old. The initial PS values for these children overlapped with the initial values seen for the children with SLI in this study. It is important to determine whether the growth trajectories are the same for the two groups of children during this early period of tense/agreement use.

The data from this study can also serve as an important downward extension of an earlier growth curve study by Rice et al. (1998). These investigators calculated FVMC scores for children with SLI from the age of 4;6 to 8;8 years;months. The ages of our participants at the end of our study overlapped with the ages of the participants at the starting point in the Rice et al. study. A comparison of the data across the studies should suggest whether trends seen in our study might be expected to continue, or change, with increasing age.

Method

Participants

Seventeen monolingual English-speaking children served as participants in the study. Data for these children originated with the intervention study initially reported by Leonard, Camarata, Brown, and Camarata (2004), with additional data provided in Leonard, Camarata, Pawłowska, Brown, and Camarata (2006, 2008). These 17 children ranged in age from 36 to 51 months (M = 42 months). Eleven children were boys, and six were girls. The parents of all children had contacted a clinical service provider about their concerns about their children's language development. Upon testing, all children scored below the 10th percentile on both the Structured Photographic Expressive Language Test–Preschool (Werner & Kresheck, 1983) and Developmental Sentence Scoring (Lee, 1974). The children passed a hearing screening and scored at least 85 (M = 105) on the Leiter International Performance Scale–Revised (Roid & Miller, 1997). In addition, all children fell well within the "nonautistic" end of the continuum on the Childhood Autism Rating Scale (Schopler, Reichler, & Renner, 1988).

The children showed no more than 20% use of the intervention target of this study—third-person singular *–s*—before treatment, based on a 12-item probe task and an inspection of their spontaneous speech sample. Eleven of the 17 children showed 0% use of third-person singular *–s*. The children met the same limited-use criteria for auxiliary *is, are,* and *was* and past tense *–ed*. Third-person singular *–s* was judged to be an appropriate target for intervention because all children already showed some degree of use of early-emerging grammatical morphemes such as progressive *–ing* and noun plural *–s*. To be sure that the children's production of third-person singular *–s* would not be impeded by phonological factors, we also examined the children's ability to produce word-final /s/ and /z/ (as well as word-final /t/ and /d/) in a probe task of monomorphemic words (e.g., *fox, nose*). All children showed at least 80% accuracy on this probe. Finally, all children exhibited a mean length of utterance sufficient to support the use of all tense/agreement morphemes studied here.

Treatment Procedure for Third-Person Singular –s

The children participated in 96 treatment sessions. In most cases, the sessions were scheduled for 2 days per week, with two sessions held on each of these days. Variations in this schedule sometimes occurred due to illness, weather-related transportation problems, or other factors. Details of the treatment procedures can be seen in Leonard et al. (2004); the posttreatment follow-up session is described in Leonard et al. (2008). Each treatment session was divided into two segments—a focused stimulation segment and a recasting segment. During the focused stimulation segment, the clinician read a story to the child and acted out the story with toys. In each story, the third-person singular *–s* inflection was used 12 times, with at least six different verbs. Once the story had been read, the recasting segment commenced. The clinician and child played with the toys used in the stories, and the clinician looked for opportunities to recast the child's utterances with a conversationally appropriate reply that contained the third-person singular *–s*. Twelve such recasts were provided during each session. During both the focused stimulation and recasting segments of the session, care was taken to avoid the use of past tense *–ed* as well as both copula and auxiliary *is, are, am, was,* and *were*. However, some of the stories did include auxiliary *do* forms.

As reported by Leonard et al. (2004, 2006, 2008), treatment fidelity was strong, and the children made gains in the use of the target third-person singular *–s* forms that seemed attributable to the treatment procedures. Growth curve analyses (GCAs) of these gains were not conducted. In addition, the composite scores of the FVMC, TMT, and PS were not calculated in the original treatment studies.

To assess the progress during and after the intervention task, the 12-item third-person singular *–s* probe task presented before intervention was repeated after the 48th intervention session (an average of 24 weeks after the preintervention administration), after the 96th intervention session (an average of 44 weeks after the preintervention administration), and approximately 7 weeks after the end of intervention (an average of 51 weeks after the preintervention administration). Hereafter, these time points will be referred to as Times 1, 2, 3, and 4, respectively. Each of the 12 items required the child to use a different verb in a habitual action context following a sentence completion format. For the 17 children studied here, the mean percentages correct for Times 1–4 were 3.71 (SD = 5.58), 28.82 (SD = 27.44), 52.12 (SD = 38.18), and 57.00 (SD = 39.11), respectively.

Spontaneous Speech Samples

A spontaneous speech sample was also obtained from each child during Times 1–4. The sample was obtained as the examiner and child played with developmentally appropriate toys. The average number of complete and intelligible utterances per sample for the four time points was 175 (range = 109–353), 221 (range = 96–379), 216 (range = 124–300), and 248 (range = 153–456), respectively.

The Composite Measures

The children's spontaneous speech samples were used to compute the three composite measures—the FVMC, the TMT, and the PS. For each of the four time points, we computed these composites from complete and intelligible utterances for each child's full sample and for the first 100 utterances of the sample. The latter allowed us to control for sample size across the four time points.

FVMC

For each sample, the child's productions of past tense *–ed*; auxiliary *is, are, am, was,* and *were*; copula *is, are, am, was,* and *were*; and auxiliary *do, does,* and *did* were summed and divided by the total obligatory contexts for these morphemes. This value was then multiplied by 100 to yield a percentage. Multiple productions of past tense *–ed* with the same lexical verb were allowed, and all contracted and uncontracted auxiliary and copula forms were permitted, whether the subject was a noun or a pronoun. Overregularizations of past tense *–ed* were counted as correct, but all other errors were excluded. Auxiliary *do* forms with negative contractions were included in the case of *doesn't* and *didn't,* but *don't* was excluded from consideration. Note that we calculated the FVMC without third-person singular *–s,* as this morpheme was the focus of intervention.

TMT

The 14 tense/agreement morphemes scored for the FVMC were also the morphemes used for the TMT. Unlike the TMT originally designed by Hadley and Short (2005), we excluded the third-person singular *–s* inflection. For this measure, morphemes contracted to pronouns (e.g., *it's*) were excluded. Uncontracted forms used with pronouns (e.g., *it is*) and both contracted and uncontracted forms used with nouns (e.g., *girl's, dog is*) were included. Overregularizations of past tense were counted, with all other types of errors excluded from consideration. The auxiliary *do* forms *does* and *did* were included even if they appeared with negative contractions (*doesn't, didn't*), but *don't* was excluded. The TMT was the total number of different tense/agreement morphemes out of the 14 that were used at least once by the child, resulting in a possible TMT range of 0–14 for each sample. Note that this represents a measure of tense/agreement morpheme diversity but is silent on the productivity of the morphemes, as each morpheme is counted only once.

PS

For this measure, third-person singular *–s* was excluded. In other respects, the PS was computed as originally designed by Hadley and Short (2005). The 14 tense/agreement morphemes noted above were divided into separate categories. Copula *is, are, am, was,* and *were* formed the Copula BE category; auxiliary *is, are, am, was,* and *were* made up the Auxiliary BE category; and auxiliary *do, does,* and *did* represented the Auxiliary DO category. Past tense *–ed* formed its own category, Past Tense –ED, with overregularizations of past treated as correct. (All other error types were excluded.) For each of these categories, up to 5 points could be awarded, with 20 representing the highest possible combined score. These points could be earned in one of two ways. First, in the case of the Past Tense –ED category, the past tense *–ed* inflection could be used with different verbs, as in *jumped, played, opened, kicked,* and *pushed*. Or, in the case of the Copula BE, Auxiliary BE, and Auxiliary DO categories, a morpheme could be used with different subject nouns or pronouns (if uncontracted in the latter case), as in *boy is, girl is, it is, mommy's,* and *Lauren's*. Points could also be earned through a combination of different morphemes used with a variety of verbs or subject nouns. For example, 5 points would be awarded for Copula BE through a child's use of *boy is, girl is, Lauren's, cows are,* and *I was*. Although only copula *is* was used with multiple subjects in this example, the single instances of *are* and *was* would also be counted, as they form part of the category Copula BE. Given that the same morpheme is given more points if it is used with different verbs/subjects, the PS provides more of an indication of a child's tense/agreement productivity than does the TMT.

Reliability

To assess scoring reliability, a second coder was trained to compute the three composite scores. This second coder then scored 25% of the samples, drawing from each of the four measurement points. The interobserver agreement was 96.31% for the FVMC, 97.90% for the TMT, and 95.00% for the PS.

Results

The three composites—the FVMC, TMT, and PS—were each calculated using two sample sizes. For the first set of analyses, we used the complete sample from each child at each time point. With this analysis, we could use the full range of data at our disposal and perhaps obtain a more representative picture of the children's abilities as estimated by each of the composites. For the second set of analyses, we selected only the first 100 spontaneous utterances from each child's sample (except for the Time 2 sample of one child, which contained 96 utterances). This set of analyses allowed us to observe the developmental trajectory of each composite type when the samples were constrained to the same length. A sample size of 100 utterances seemed especially appropriate given that it is probably the maximum sample size to be employed by clinicians during their assessment of children's language ability. Because the TMT and PS are additive measures, it seemed likely that the scores for these measures would be appreciably lower at 100 utterances than for the full sample. However, the developmental trajectories for the two sample sizes cannot be predicted based on sample size alone.

Descriptive Data

Table 1 provides the mean composite scores for the FVMC, TMT, and PS at each time point, for each sample size. Although the mean values for all of the composites suggest at least minimal use at Time 1, two children showed 0% use on the FVMC; these two children as well as a third child earned no points on both the TMT and PS composites.

As can be seen from Table 1, scores at the group level increased across time on all measures, especially from Times 1 to 3. This is the period when the children were participating in the intervention sessions. The smallest gains occurred from Times 3 to 4. This was a period after intervention, but it might also be significant that a shorter amount of time had transpired between these two time points than between each of the other time points. This shorter period is taken into account in our GCAs reported below.

GCAs

To address our research questions that asked how the composite scores of the FVMC, TMT, and PS developed in children with SLI during the intervention phase and after intervention, we used mixed-effects GCAs (Mirman, 2014). We modeled the growth trajectory of each composite score over the four time points in separate models. Therefore, the dependent variables in each model were the composite scores of the FVMC, TMT, or PS. As in previous studies examining the trajectory of tense/agreement morphemes (Rice et al., 1998; Rispoli, Hadley, & Holt, 2009), the fixed-effect predictor variables included linear and quadratic time (measured in weeks from the first assessment for each participant's Time 1 session). Before entering this information into the mixed-effect regression, we divided each time value by the largest time elapsed between the first and final sessions to facilitate convergence in our models, which obtain coefficients via maximum likelihood optimization. The intercept was set at Time 1 because we were interested in capturing the composite score growth throughout the intervention and the maintenance or continued growth after the intervention ended. In addition, all models included random effects at the participant level across time terms. Last, we used the z distribution to evaluate significance ($t > \pm 1.95$ was considered significant at the .05 level). Models were run using the R package lme4 (Bates, Maechler, Bolker, & Walker, 2014).

FVMC Growth

Figure 1 depicts the growth trajectories of the FVMC across the four time points, for both the full samples and the 100-utterance samples.

Full samples. A random-slope GCA model was used to examine the growth of the overall increase in the FVMC scores. There was a significant effect of intercept and linear time (estimate = 17.37, SE = 3.99, t = 4.36; and estimate = 70.48, SE = 18.26, t = 3.86, respectively), indicating that FVMC scores were emerging at Time 1 and continued to grow throughout the intervention. In addition, there was a significant effect of quadratic time (estimate = −35.05, SE = 17.73, t = −1.98), indicating that the rate of growth of FVMC scores over time decreased.

To determine whether the deceleration of FVMC growth occurred during the intervention or after the intervention ended, we conducted a follow-up, random-slope GCA model that included the FVMC scores from Times 1 through 3, thus excluding the follow-up assessment after the intervention had ended at Time 4. As before, there was a significant effect of intercept and linear time (estimate = 16.84, SE = 5.13, t = 3.28; and estimate = 86.94, SE = 23.19, t = 3.75, respectively), indicating that FVMC scores were significantly higher than zero at Time 1 and grew during the intervention. There also was a significant effect of quadratic time (estimate = −60.86, SE = 30.42, t = −2.00), indicating that growth of FVMC scores decelerated toward the end of the intervention.

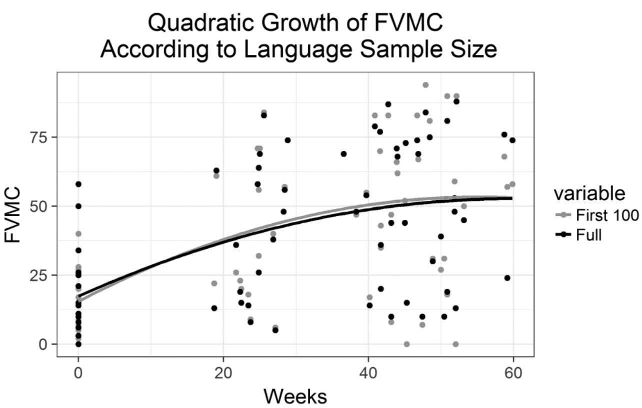
The FVMC differs from the other two composites in its inclusion of tense/agreement morphemes that are used in contracted form with pronoun subjects. This inclusion may have contributed to the pattern of growth seen for the FVMC, because tense/agreement morphemes in these contexts are frequent in the input and may be produced at an earlier point in time by children through direct activation. After this initial surge, growth might decelerate as

Table 1. Growth of tense/agreement composites over time.

Measure	Time 1	Time 2	Time 3	Time 4
FVMC-full	17.29 (16.84)	40.53 (25.65)	48.65 (22.71)	52.65 (30.33)
FVMC-100	15.29 (16.08)	42.24 (24.82)	50.12 (21.01)	53.94 (32.44)
TMT-full	1.76 (1.60)	3.65 (2.40)	4.82 (2.43)	5.12 (2.60)
TMT-100	1.18 (0.95)	1.76 (1.56)	3.12 (1.73)	3.24 (2.28)
PS-full	2.35 (2.55)	6.06 (4.72)	8.76 (4.22)	8.76 (4.78)
PS-100	1.41 (1.28)	2.71 (2.62)	4.82 (2.86)	5.00 (3.94)

Note. Mean scores (and standard deviations) for each composite measure and the morpheme targeted for treatment, third-person singular *–s*, across the four time points. FVMC = Finite Verb Morphology Composite (maximum score = 100); TMT = Tense Marker Total (maximum score = 14); PS = Productivity Score (maximum score = 20); full = value for a complete spontaneous speech sample; 100 = value for a 100-utterance sample.

Figure 1. Quadratic growth of the Finite Verb Morphology Composite (FVMC) for the 100-utterance and full samples.



changes in the FVMC might begin to depend more heavily on grammatical encoding.

To test the plausibility that gains more likely to be attributable to direct activation reflect a different pattern from gains more likely to be attributable to grammatical encoding, we examined each child's use of copula *is*. This morpheme is by far the most frequent both in its obligatory contexts and in its percentage of use in these contexts. In Figure 2, we show the children's percentage of the use of copula *is* in obligatory contexts from Times 1 to 3 when it was used in contracted form with a pronoun subject (indexing use via direct activation) and in all other contexts (indexing use via grammatical encoding). As can be seen in the figure, there is an especially steep gain in the contraction-with-pronoun forms from Times 1 to 2; thereafter, gains are shallower. In contrast, gains in the use of copula *is* in all other contexts combined appear to be rather consistent across the time points. However, the slightly higher percentage in obligatory contexts for the contraction-with-pronoun forms at Time 2 obscures the fact that these forms were produced much more frequently than copula *is* in other contexts, as can be seen in Figure 3. Given the manner in which the FVMC is computed, this much higher frequency of contraction-with-pronoun forms, coupled with their slightly higher percentage of use in obligatory contexts, probably had an especially strong influence on the steep gains from Times 1 to 2 on this composite.

In contrast, the percentage of use of contraction-with-pronoun forms increased only slightly from Times 2 to 3 (see Figure 2). Although the actual number of productions of these forms declined somewhat (because of fewer obligatory contexts) relative to Time 2, the number of these productions was still substantially higher than those of copula *is* in the other contexts (see Figure 3). This suggests that the smaller gains from Times 2 to 3 in the contraction-with-pronoun forms probably had a greater influence on the trajectory of the FVMC than the larger gains seen for copula *is* in the other contexts.

100-utterance samples. As with the full samples, a random-slope GCA model was used to examine the growth of the overall increase in the FVMC scores for the 100-utterance samples. There was a significant effect of intercept and linear time (estimate = 15.48, SE = 3.81, t = 4.06; and estimate = 82.23, SE = 20.97, t = 3.92, respectively), indicating that FVMC scores were emerging at Time 1 and continued to grow throughout the intervention. In addition, there was a significant effect of quadratic time (estimate = −44.57, SE = 21.86, t = −2.04), indicating that the rate of growth of FVMC scores over time decreased.

To determine whether the deceleration of FVMC growth occurred during the intervention or after the intervention ended, we conducted a follow-up, random-slope GCA model that included the FVMC scores from Times 1 through 3, excluding the follow-up assessment after the intervention had ended at Time 4. As in the model with all four time points, there was a significant effect of intercept and linear time (estimate = 14.94, SE = 4.07, t = 3.67;

Figure 2. Mean percentage correct of copula *is* in obligatory contexts.

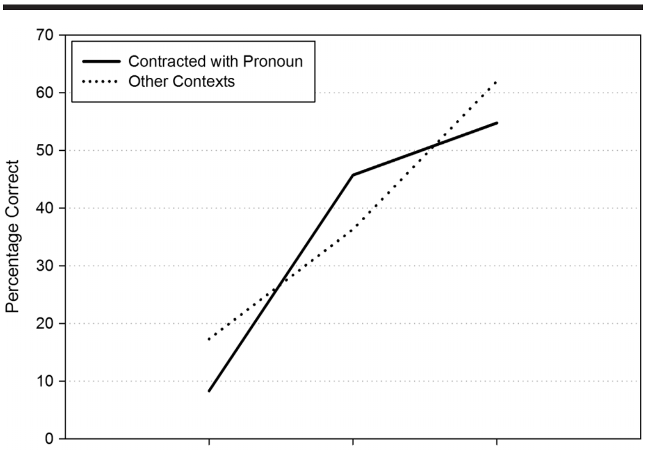
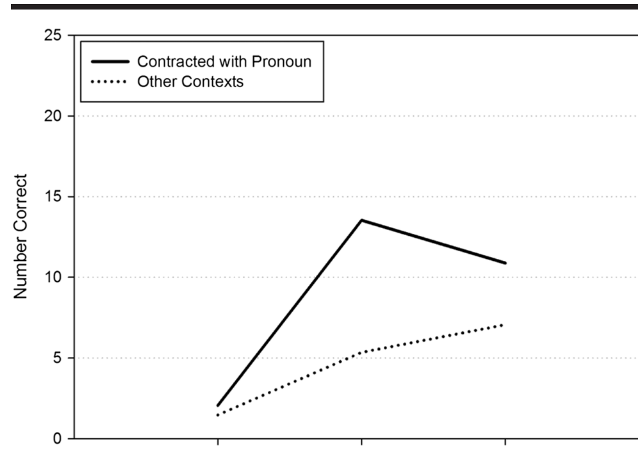


Figure 3. Mean number of correct productions of copula *is*.



and estimate = 96.49, SE = 23.24, t = 4.15, respectively), indicating that FVMC scores were significantly higher than zero at Time 1 and grew during the intervention. There also was a significant effect of quadratic time (estimate = −66.61, SE = 30.32, t = −2.20). The latter finding indicated that, even with samples of 100 utterances, there was a detectable deceleration of growth of FVMC scores toward the end of the intervention.

From an inspection of Figure 1, it is clear that the trajectories for the FVMC with the two sample sizes are striking in their similarity. Both seem to capture the same pattern of development, despite the fact that the full samples are not only larger than the 100-utterance samples but also increase in size across the four time points (e.g., means of 221 and 248 at Time 2 and Time 4, respectively).

TMT Growth

Figure 4 depicts the longitudinal growth of TMT across the four time points for both the full samples and the samples of 100 utterances.

Full samples. The random-slope GCA model predicting the growth of TMT scores revealed a significant effect of intercept and linear time (estimate = 1.76, SE = 0.42, t = 4.24; and estimate = 5.41, SE = 2.41, t = 2.26, respectively), indicating that TMT scores were significantly greater than 0 at Time 1 and continued to grow throughout the intervention. However, quadratic time was not significant. These results indicate that the growth of TMT scores is best described as having linear growth across the four time points.

100-utterance samples. The random-slope GCA model predicting the growth of TMT scores revealed a significant effect of intercept (estimate = 1.18, SE = 0.30, t = 3.88), indicating that TMT scores derived from the first 100 utterances were significantly greater than 0 at Time 1. Neither linear nor quadratic growth was observed (estimate = 0.78, SE = 1.86, t = 0.42; and estimate = 2.07, SE = 2.19, t = 0.95, respectively).

After visual inspection of the data, we tested an exploratory random-slope GCA model with only intercept and linear time terms as fixed effects. Scores at Time 1 (intercept) and linear growth were allowed to vary (i.e., random effects). In this simplified model, there were significant effects of both intercept and linear time (estimate = 1.05, SE = 0.30, t = 3.51; and estimate = 2.59, SE = 0.61, t = 4.26, respectively), indicating that TMT scores were greater than 0 at Time 1 and grew across the four time points.

Although both of the trajectories shown in Figure 4 appear to reflect linear growth, the trajectory for the full samples shows a steeper slope than for the 100-utterance samples. This difference probably reflects sample size differences, such that the 100-utterance samples underestimated the number of different tense/agreement morphemes that the children could produce at least to a minimal degree.

PS Growth

Figure 5 provides the developmental trajectories for the PS at all four time points for both sample sizes.

Full samples. A random-slope GCA model examined the growth of PS. There was a significant effect of intercept and linear time (estimate = 2.32, SE = 0.65, t = 3.58; and estimate = 11.34, SE = 4.97, t = 2.28, respectively) but not quadratic time. Growth of PS across the four time points was best described as a linear slope.

100-utterance samples. A random-slope GCA model also examined the growth of the PS for the 100-utterance samples. The model revealed a significant effect of intercept (estimate = 1.42, SE = 0.43, t = 3.28); however, both linear growth (estimate = 2.67, SE = 2.96, t = 0.90) and quadratic growth (estimate = 1.96, SE = 3.35, t = 0.59) failed to reach significance.

A follow-up analysis further explored changes in the PS across time. After visual inspection of the data, we used a random-slope GCA model that contained intercept and linear time terms as fixed effects. Participants were also

Figure 4. Linear growth of the Tense Marker Total (TMT) for the 100-utterance and full samples.

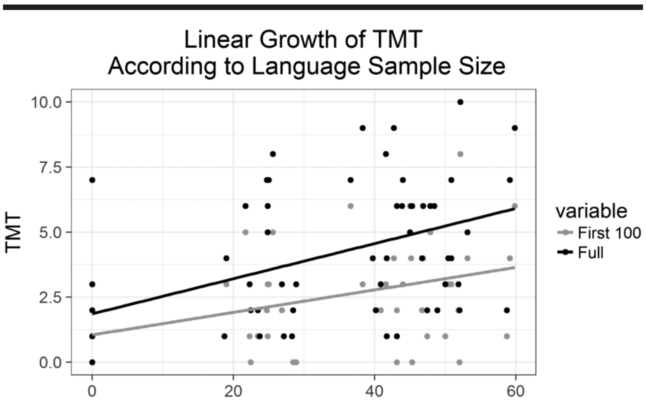
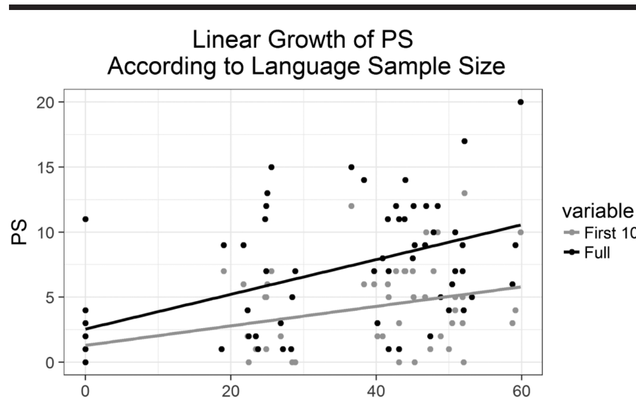


Figure 5. Linear growth of the Productivity Score (PS) for the 100-utterance and full samples.



allowed to vary at Time 1 and were allowed to vary in linear growth (random effects). In this simplified model, there was a significant effect for both intercept and linear time (estimate = 1.29, *SE* = 0.43, *t* = 3.00; and estimate = 4.50, *SE* = 1.12, *t* = 4.03, respectively), indicating that children's PS composites were greater than zero at Time 1 and grew across the four time points.

An inspection of the two trajectories in Figure 5 shows somewhat similar slopes for the two sample sizes, although it appears that the larger sample size provided the children with more opportunity to show productivity than was true for the samples of 100 utterances.

Although the PS composites showed steady growth even with samples of 100 utterances, we took an extra step to determine whether the scores from these smaller samples were representative in their composition. Following Rispoli et al. (2012), we computed the PS separately for each of the tense/agreement categories used for this composite, excluding the third-person singular *–s,* as this was the target in the original intervention study. On the basis of these scores, we rank-ordered the categories (1 = highest score) for each time point and compared these rank orders with those reported by Rispoli et al. for 33 months (the final time point in their study) and those of Gladfelter and Leonard (2013) for both their typically developing group (*M* age = 57 months) and for their group with SLI (*M* age = 57 months). The ranks are shown in Table 2. As can be seen, in the earlier studies, Copula BE consistently showed the highest scores, followed by Auxiliary DO. Past –ED and Auxiliary BE had lower scores, with a rank of either 3 or 4, depending on the study. It can also be seen that at each time point in this study, the same pattern emerged. Regardless of time point, Copula BE scores were highest, followed by Auxiliary DO scores, with Past –ED and Auxiliary BE always having a rank of either 3 or 4. At each time point, approximately half of the Auxiliary DO productions were *doesn't* or *didn't* in declarative contexts; auxiliary-fronted questions with *do* were usually affirmative (nonnegative forms). Note also that the PS values for Copula BE do not include copula forms contracted with pronoun subjects. Therefore, although Copula BE showed the highest scores among the tense/agreement categories of the PS, these scores were independent of the highly frequent pronoun-with-contraction forms responsible for the early and steep increase in the FVMC.

Table 2. Rank order of the productivity scores for each tense/agreement category (1 = highest score) at each time point in this study and in the studies of Rispoli et al. (2012) and Gladfelter et al. (2013).

Study	Copula BE	Auxiliary DO	Past –ED	Auxiliary BE
This study				
Time 1: 42 mos	1	2	4	3
Time 2: 48 mos	1	2	3	4
Time 3: 53 mos	1	2	3	4
Time 4: 55 mos	1	2	4	3
Rispoli et al.				
TD 33 mos	1	2	3	4
Gladfelter et al.				
TD 57 mos	1	2	4	3
SLI 57 mos	1	2	3	4

Note. mos = age in months; SLI = children with specific language impairment; TD = children with typical language development.

Discussion

The three tense/agreement composites studied here were selected because they have been shown to be both developmentally sensitive (Goffman & Leonard, 2000; Rispoli et al., 2012) and diagnostically accurate (Gladfelter & Leonard, 2013). All three measures increase from the earliest stages of tense/agreement development through the preschool years, and each distinguishes children with SLI from their typically developing peers. Yet, despite these shared characteristics, these measures differ in important ways. Part of this difference lies in their relative emphasis on accuracy, diversity, or productivity. However, the manner in which accuracy, diversity, and productivity are operationalized in these measures introduces another kind of difference. Diversity as measured by the TMT and productivity as indexed by the PS are additive and, as a result, are influenced to a greater degree by sample size than is the case for the FVMC, the measure of accuracy. In discussing our findings for the three composites, we point out characteristics of their developmental trajectories that appear to reflect their relative emphasis (on accuracy, diversity, or productivity) as well as characteristics that seem to reflect the influence of the size of the spontaneous speech sample on which the composites were based.

We had expected the growth trajectories of the three composites to differ, and this appears to be the case. For analyses focused on all four time points, the slopes of all three composites were significant for both intercept and linear time. In other words, despite the children's clear tense/agreement limitations, their composite scores differed from zero before the intervention began, and growth continued across the four time points. However, only the FVMC revealed a developmental trajectory of quadratic and linear growth. As was seen in Figure 1, the shallower slope at a later time point might have been due to a reduction in the rate of change once treatment ended at Time 3. Although the FVMC was calculated without the third-person singular *–s* and clinicians avoided the use of the most frequent of the remaining tense/agreement morphemes, growth could have slowed due to less overall language stimulation from Times 3 to 4. However, analysis of the growth of the FVMC from Times 1 through 3 only also revealed a significant quadratic function, suggesting that the rate of growth was slowing before the cessation of treatment.

One possible explanation of this profile is that the early period of steeper growth was influenced by the children's rapid accumulation of frequently occurring forms. Our observation of steeper growth from Times 1 to 2 for copula *is* in contraction-with-pronoun contexts than in other contexts seems consistent with this possibility. Note that although contractions with pronouns could be the result of direct activation (Rispoli et al., 2012), it seems to be the sheer frequency of their use by the children that shaped the growth curve for the FVMC, not their specific grammatical status in the children's grammars. That is, even if one assumes that these forms were generated in the same way as tense/agreement forms in other contexts, much of the early growth in overall tense/agreement accuracy could have been driven by their high frequency.

From the results of the GCAs and from Figure 1, it is clear that changes in the FVMC are not influenced greatly by sample size, at least when a minimum of 100 spontaneous utterances are used. The early influence of contraction-with-pronoun forms is present even with the smaller sample size. Just as notable is that a deceleration is seen in the later samples, regardless of sample size. Note, for example, that the deceleration in the growth from Times 2 to 4 for the full samples occurred, although an average of 248 utterances were used at Time 4, but an average of only 221 utterances were employed at Time 2. Furthermore, given that the children's accuracy as measured by the FVMC averaged under 60% even at Time 4, the deceleration cannot be attributed to ceiling effects.

The seemingly robust decelerating trajectory of the FVMC serves as an important guide to understanding the course of tense/agreement accuracy, especially when overall accuracy is of interest rather than the pattern of individual morphemes. Specifically, an early surge in growth might be expected, thanks in large part to children's early development of tense/agreement morphemes that appear in frequently occurring contexts. The pace of this development will begin to slow at the same time when tense/agreement morphemes in less frequently occurring contexts show steady gains in accuracy. However, the latter pattern may be masked because the forms whose gains are decelerating are nevertheless more frequent in occurrence, thus exerting a greater influence on the FVMC.

The FVMC scores of this study can be compared with those reported by Rice et al. (1998) for children with SLI from ages 4;6 to 8;8 years;months. Recall that in our study, a deceleration was apparent from Times 2 to 3, when the children's FVMC scores changed, averaging approximately from 40% to 50% (for both full and 100-utterance samples). Rice et al. found a similar rate of change until the children averaged approximately 75% at the age of 6;6 years;months, at which point a more striking deceleration was observed. In fact, the children's scores did not reach asymptote even by the end of the study. Although extrapolations should be interpreted with caution, the Rice et al. findings suggest that the quadratic trajectory in the FVMC scores in our study is not uncharacteristic of the patterns expected for children with SLI.

Both the TMT and PS composites were sensitive to growth over the period of the study. This was seen most clearly in the growth trajectories for the 100-utterance samples, where sample size was controlled. Because these two measures exclude contractions of tense/agreement morphemes with pronouns, we did not expect them to show the steep gains seen for the FVMC from Times 1 to 2. The TMT and PS composites also differed from the FVMC in being more affected by sample size. This effect is likely attributable to the additive nature of these measures.

Earlier, we noted that the interpretation of contractions with pronouns as reflecting direct activation or grammatical encoding is not essential to the FVMC. For that measure, the importance lies in the high frequency and early appearance of these forms. For the TMT and PS, on the other hand, the possibility that contraction-with-pronoun forms are the result of direct activation was the basis for their exclusion (Hadley & Short, 2005). We adhered to the same exclusion criteria in this study for two reasons. First, the TMT and PS, as defined by Hadley and Short, have already produced valuable developmental and clinical data and thus seemed appropriate for further study in its current form. Second, even if a basis for the inclusion of these forms were devised (e.g., if used with multiple pronouns, such as *it's, that's, he's,* and *she's*), their high frequency and early appearance would have a disproportionate influence on these measures, rendering them more like the FVMC. As a result, any appearance of forms most likely to reflect grammatical encoding (e.g., *monkey's, fish is*) might be missed because the single-appearance criterion of the TMT and the five-different-context criterion of the PS will have already been met by the more frequently produced contractions with pronoun subjects.

The observed characteristics of the TMT and PS composites have both positive and negative implications. Regarding the former, the linear growth seen in these measures even when sample size is held constant at 100 utterances means that these measures might serve as useful gauges of change in tense/agreement diversity and productivity by children with SLI during the developmental period studied here. Note in this context that the 100-utterance samples at each time point reflected the same rank orders of tense/agreement category use in the PS that have been reported in earlier studies with larger sample sizes. Although PS values would no doubt rise with increasing sample size, we believe that the common sequence seen across studies is due in part to the order in which children learn the surface morphemes that correspond to the tense and agreement features they are beginning to grasp.

The drawback is that the effects of sample size on these measures mean that children's TMT and PS composites cannot be interpreted in normative terms. For example, a difference in PS composites between, say, 5.00 and 8.76 might be taken as large and clinically significant. Yet, these were the two means for the Time 4 100-utterance and full samples from the same children in this study. Strict standardization of sample size will be necessary before these additive measures can be used in this way.

We are struck by the differences between the growth of the PS in our data and the growth seen in the young typically developing children studied by Rispoli et al. (2009) between the ages of 21 and 30 months. We observed only linear change, whereas Rispoli et al. found accelerating quadratic growth. The PS means and standard deviations for the typically developing children at 24 and 27 months old were very similar to those in our study at Time 1 (when our participants averaged 42 months old) and Time 2 (when average age was 48 months), respectively. However, the acceleration was quite apparent from 27 to 30 months old in the typically developing children and decidedly absent from Times 2 to 3 (when average age was 53 months) in our participants.

We are unable to find a plausible explanation for this difference between the two studies apart from the fact that one group exhibited SLI and the other appeared to be developing language in a typical manner. We do not believe that sample size differences can explain the divergent findings. One-hour speech samples were used by Rispoli et al. (2009), and these were larger than the samples we employed. Because larger sample sizes provide more opportunities for children to earn points on the PS, the similar scores for the two groups of children suggest that the participants in the Rispoli et al. study were no more and possibly even less advanced than our participants at the first two time points. The subsequent time point revealed accelerating growth for the Rispoli et al. participants, despite a sample size similar to those used at the first two time points. On the other hand, no such acceleration was seen in our participants, although the samples in question were slightly larger than the samples at the previous time points. The difference between the two studies also does not seem attributable to our exclusion of third-person singular *–s* from the calculation of PS because this morpheme already influences this measure at earlier ages (Rispoli et al., 2012).

An explanation for the difference that seems more plausible is that whatever biologically driven process promoted an acceleration in the typically developing children had simply not yet occurred in the children with SLI studied here. A major tenet of the "extended optional infinitive" hypothesis of Rice and her colleagues is that the resolution of the inconsistency in the use of tense/agreement morphemes by children with SLI occurs at a significantly greater age than is seen in typical development (Rice & Wexler, 1996; Rice et al., 1998). Once this point of resolution is reached, an acceleration might be seen in these children's PS composites as well. This point had not been reached by our participants with SLI by the end of the study, a conclusion that seems reinforced by their still low scores on all three composites, as seen in Table 1.

As time progressed, even with a uniform sample size of 100 utterances, growth appeared to be shallower for the FVMC than for the TMT and the PS. This suggests that increases in the TMT and the PS might have occurred even with minimal or no change in the consistency with which a child used tense/agreement morphemes. This is especially clear in the case of the TMT, for it requires the production of only a single past tense inflection or a single copula or auxiliary form to produce a demonstrable increase in this composite, with very little effect on the FVMC. The PS, too, can show a measurable change without a corresponding change in the FVMC. For example, even if children are in a period where there is little change in their overall level of tense/agreement consistency, they might learn new verbs that are then inflected (albeit inconsistently) and new nouns that can serve as subjects (with which copula and auxiliary forms can be combined, again inconsistently).

Of course, because there is a ceiling above which the PS cannot increase further, there can reach a point when this measure will no longer be sensitive to developmental change. Even if the FVMC continues to increase at that point, one cannot assume that consistency is catching up to productivity; additional changes in productivity might simply be beyond the reach of the PS. Recall that the PS, like the TMT, was designed to capture the early signs of tense/agreement growth (Hadley & Holt, 2006; Hadley & Short, 2005).

However, this constraint does not appear to be operating at the levels of development studied here, and we regard this as a very important finding. It suggests that, independent of sample size, increased diversity and productivity might have a minimal (or no) effect on overall consistency. Furthermore, this increase might be missed if the focus is strictly on children's percentage of use in obligatory contexts. Although normative use of the TMT and PS will require strict standardization of sample size, it is already clear that information about children's development of tense/agreement will be incomplete without supplementing accuracy measures with those designed to estimate children's diversity and productivity in using these morphemes.

Acknowledgments

The work reported in this article was supported in part by Grant R01DC04544 from NIDCD of the National Institutes of Health (awarded to Laurence B. Leonard).

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed effects models using Eigen and S4 (R package version 1.1-7)*. Retrieved from http://CRAN.Rproject.org/package=lme4

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

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

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

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.

Hadley, P., & Holt, J. (2006). Individual differences in the onset of tense marking: A growth-curve analysis. *Journal of Speech, Language, and Hearing Research, 49,* 984–1000.

Hadley, P., & Short, H. (2005). The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48,* 1344–1362.

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

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

Leonard, L. (2014b). Replacing one imperfect term with another. *International Journal of Language and Communication Disorders, 49,* 436–437.

Leonard, L., Camarata, S., Brown, B., & Camarata, M. (2004). Tense and agreement in the speech of children with specific language impairment: Patterns of generalization through intervention. *Journal of Speech, Language, and Hearing Research, 47,* 1363–1379.

Leonard, L., Camarata, S., Pawłowska, M., Brown, B., & Camarata, M. (2006). Tense and agreement morphemes in the speech of children with specific language impairment during intervention: Phase II. *Journal of Speech, Language, and Hearing Research, 49,* 749–770.

Leonard, L., Camarata, S., Pawłowska, M., Brown, B., & Camarata, M. (2008). The acquisition of tense and agreement morphemes by children with specific language impairment during intervention: Phase 3. *Journal of Speech, Language, and Hearing Research, 51,* 120–125.

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

Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press.

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

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

Rice, M., 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.

Rispoli, M., Hadley, P., & Holt, J. (2009). The growth of tense productivity. *Journal of Speech, Language, and Hearing Research, 52,* 930–944.

Rispoli, M., Hadley, P., & Holt, J. (2012). Sequence and system in the acquisition of tense and agreement. *Journal of Speech, Language, and Hearing Research, 55,* 1007–1021.

Roid, G., & Miller, L. (1997). *Leiter International Performance Scale–Revised*. Wood Dale, IL: Stoelting.

Schopler, E., Reichler, R., & Renner, B. R. (1988). *The Childhood Autism Rating Scale*. Los Angeles, CA: Western Psychological Services.

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

Werner, E., & Kresheck, J. (1983). *Structural Photographic Expressive Language Test–Preschool*. DeKalb, IL: Janelle.