



Published in final edited form as:

J Speech Lang Hear Res. 2013 April ; 56(2): 542–552. doi:10.1044/1092-4388(2012/12-0100).

Alternative Tense and Agreement Morpheme Measures for Assessing Grammatical Deficits During the Preschool Period

Allison Gladfelter and Laurence B. Leonard
Purdue University

Abstract

Purpose—Hadley and Short (2005) developed a set of measures designed to assess the emerging diversity and productivity of tense and agreement (T/A) morpheme use by two-year-olds. We extend two of these measures to the preschool years to evaluate their utility in distinguishing children with specific language impairment (SLI) from their typically developing (TD) peers.

Method—Spontaneous speech samples from 55 children (25 children with SLI, 30 TD children) at two different age levels (4;0–4;6, 5;0–5;6) were analyzed, using a traditional T/A morphology composite that assessed accuracy, and the Hadley and Short measures of Tense Marker Total (assessing diversity of T/A morpheme use) and Productivity Score (assessing productivity of major T/A categories).

Results—All three measures showed acceptable levels of sensitivity and specificity. In addition, similar differences in levels of productivity across T/A categories were seen in the TD and SLI groups.

Conclusions—The Tense Marker Total and Productivity Score measures seem to have considerable utility for preschool-aged children, by providing information about specific T/A morphemes and major T/A categories that are not distinguished using the traditional composite measure. The findings are discussed within the framework of the Gradual Morphosyntactic Learning account.

Introduction

One of the most distinguishing features of English-speaking children with specific language impairment (SLI) is their extraordinary difficulty in the use of grammatical morphemes that reflect tense and agreement (T/A). These morphemes include third person singular *-s*, past tense *-ed*, the auxiliary *do* forms (*do, does, did*), and both finite copula and auxiliary *be* forms (*is, are, am, was, were*). The children's difficulty in the use of these morphemes stands out from many of their other language difficulties in two ways. First, during the preschool years, children with SLI make less use of these morphemes than younger typically developing children matched for mean length of utterance (MLU) (Rice, Wexler, & Cleave, 1995) and verb diversity (Leonard, Miller, & Gerber, 1999). This is noteworthy because the MLU matching ensures that utterance length or verb inventory limitations are not responsible for the T/A difficulty.

Second, during the preschool and early school ages, the use of T/A morphemes shows good diagnostic accuracy in distinguishing children with SLI from their typically developing peers. This is seen especially clearly when composite measures of T/A use are employed. These composite measures have been based on the children's spontaneous speech samples.

For example, Bedore and Leonard (1998) computed a finite verb morphology composite by noting all appropriate productions of third person singular *-s*, past tense *-ed*, and both the finite copula and auxiliary *be* forms, and then calculating the percentage of use relative to all obligatory contexts for these morphemes. Rice, Wexler, and Hershberger (1998) employed a similar composite but also included the auxiliary *do* forms. More recently, Rice and Wexler (2001) employed a subset of these morphemes in a more formal elicitation task to arrive at a similar composite. Each of these composite measures has demonstrated acceptable levels of diagnostic accuracy, based on the criterion of at least 80% for both sensitivity and specificity suggested by Plante and Vance (1994).

According to Rice and Wexler (1996), the problem with T/A use resides in children's failure to grasp that tense and agreement are obligatory in main clauses. This failure is attributed to a biologically-based principle that does not emerge until children approach three years of age. However, children with SLI are very late in acquiring this principle. As a result, they remain in a period of using T/A morphemes inconsistently for an extended duration.

There is great value in assessing T/A morpheme use in spontaneous speech because it is likely to be representative of the child's typical manner of speaking. However, this important advantage of using spontaneous speech to assess T/A morpheme use is accompanied by some drawbacks. First, composites are heavily influenced by the child's facility or weakness in using those particular T/A morphemes that are most likely to have the most obligatory contexts. For example, copula *is* contexts appear to be more frequent than past tense *-ed* contexts. Therefore, if a child shows limited or no use of past tense *-ed* in the few obligatory contexts that occur for this morpheme, and uses copula *is* with greater proficiency, a composite based on all morphemes combined could mask the weakness with past tense *-ed*.

A second drawback is that composite T/A measures based on spontaneous speech are likely to include instances of T/A morpheme use that may constitute unanalyzed wholes. For example, contracted copula *is* forms in utterances such as *It's pretty*, *That's yucky*, and *He's funny* may have been learned directly from the input with little grammatical analysis on the child's part. Such utterances need not reflect immaturity. For example, even in adult use, there are forms that co-occur so frequently that they become automated, even leading to grammatical errors (e.g., the frequent use by adults of utterances such as *There's two free seats in the second row* instead of *There are two free seats in the second row*). However, in adult use, larger units such as *That's* and *He's* exist in parallel with forms generated by grammatical rules. The problem is that, in children, it is not always clear when an utterance such as *That's yucky* or *He's funny* is still in unanalyzed form or has been generated by a productive grammatical process.

In their recent proposal of grammatical acquisition, termed the Gradual Morphosyntactic Learning account, Rispoli, Hadley, and Holt (in press) refer to these two types of utterances as reflecting direct activation and grammatical encoding, respectively. The former are utterances attributed to the influence of frequently co-occurring word-morpheme combinations that are acquired directly from the input. The latter are utterances that seem to be more clearly constructed from grammatical representations. According to Rispoli et al., forms reflecting direct activation can lead to grammatical encoding but cannot be safely attributed to grammatical processes in their initial state.

This same distinction has been incorporated into a set of measures recently developed by Hadley and her colleagues (Hadley & Holt, 2006; Hadley & Short, 2005) for assessing T/A morpheme use in the spontaneous speech of very young children. These investigators focused on the early use of T/A morphemes that seemed to reflect grammatical encoding,

and sought to avoid forms that might be attributable to direct activation. Toward this end, all utterances with contracted copula and auxiliary forms with pronoun subjects were excluded. Thus, utterances such as *She is here* and *The car's red* were included, but those such as *She's here* and *It's red* were not.

In light of the fact that inflected verb forms (e.g., third person singular *-s* in *It goes like this*) may also reflect direct activation, Hadley and colleagues also considered the diversity of lexical verbs that appeared with T/A inflections (Hadley & Holt, 2006; Hadley & Short, 2005). Greater credit was given for productions of multiple verbs with an inflection (e.g., one instance each of *goes*, *fits*, *runs*) than for multiple productions of the same inflected verb (e.g., three instances of *goes*).

Hadley and Holt (2006) found that this set of measures exhibits a clear developmental trajectory of T/A morpheme use. Furthermore, Hadley and Short (2005) found that young children's scores on these measures at 24–29 months were relatively successful in identifying children deemed at risk for SLI at age 3.

Given the advantages of these measures in emphasizing seemingly productive, grammatically generated T/A morpheme use and reducing the influence of utterances resulting from direct activation, it seemed important to determine if these measures continue to have clinical utility at an age when SLI is more definitively diagnosed. Accordingly, a major goal of the present study was to assess the diagnostic accuracy of these measures in four- and five-year-old children relative to that seen with the more traditional T/A morpheme composite that credits children with T/A morpheme use regardless of the utterances in which these morphemes appear.

We singled out for use two particular T/A measures developed by Hadley and Short (2005) because they provided a greater range of scores, thus reducing the risk of ceiling effects. The first was the Tense Marker Total (TMT). For this measure, the child is given a point for the use of each of the 15 T/A morphemes, copula *is*, *are*, *am*, *was*, *were*, auxiliary *is*, *are*, *am*, *was*, *were*, auxiliary *do*, *does*, *did*, third person singular *-s*, and past tense *-ed*. The second measure was the Productivity Score (PS). For this measure, a point is credited for up to five distinct uses of each of five T/A categories (capitalized here to distinguish them from the individual morphemes: finite COPULA BE; finite AUXILIARY BE; AUXILIARY DO, THIRD PERSON SINGULAR *-S*; and PAST TENSE *-ED*). Distinct uses can take several different forms. For a function word category (e.g., AUXILIARY BE), distinct uses can be either the same morpheme used with different subjects (e.g., *Mommy's eating*, *Heather's running*, *The girl's riding a bike*) or the use of different morphemes within the category (e.g., *Mommy's eating*, *The cats are playing*, *She was laughing at me*) or some combination of the two (e.g., *Mommy's eating*, *Heather's running*, *The cats are playing*). For an inflection category (e.g., PAST TENSE *-ED*), distinct uses are the appearance of the inflection with different verbs (e.g., *pushed*, *jumped*, *played*).

A secondary goal of the present study was to determine whether the productivity of particular T/A morphemes relative to others seen at a younger age by Rispoli et al. (in press) continues to hold for four- and five-year-olds. According to the Gradual Morphosyntactic Learning account of Rispoli et al., although T/A morphemes form a coherent constellation, they do not emerge and develop simultaneously. Three factors appear responsible for the approximate sequence in which these morphemes appear. First, T/A morphemes that are especially frequent in the input seem to have a head start. According to Rispoli et al., copula *is* will usually emerge first in children's speech, initially as an unanalyzed form and then as a form incorporated into the grammar. A second factor is the feature composition of the T/A morphemes. Given the early acquisition of copula *is*, other T/A morphemes sharing tense

and agreement features with copula *is* (present tense, third person, singular) are likely to develop soon thereafter. A third factor is the type of sentence frame that must accommodate the T/A morpheme. THIRD PERSON SINGULAR –S and PAST TENSE –ED employ the same sentence frame – a frame that lacks a slot for an auxiliary, with attachment of the T/A inflection to a lexical verb (e.g., *plays, played*). Finite COPULA forms occupy frames containing no lexical verb, whereas finite AUXILIARY BE forms appear in frames alongside a lexical verb that is inflected with progressive –*ing*. Finally, AUXILIARY DO forms are used in frames containing bare lexical verbs. Rispoli et al. found that this interaction of frequency of occurrence, shared feature composition, and distinctiveness of sentence frames led to children’s relatively early productivity of COPULA BE forms and relatively late productivity of AUXILIARY BE forms, with the remaining T/A categories (THIRD PERSON SINGULAR –S, PAST TENSE –ED, AUXILIARY DO forms) showing productivity at approximately the same time, after the COPULA BE and before the AUXILIARY BE category.

The particular productivity measure used by Rispoli et al. (in press) was the previously described Productivity Score, though with each of the five T/A categories taken separately (range = 0 to 5 for each category). At the oldest age examined, 33 months, the typically developing children in the study were approaching ceiling level for COPULA BE forms ($M = 4.11$), but were well below this level for the remaining categories (M s = 1.42 to 2.74). In the present study, we determine if a similar difference among the T/A categories in Productivity Score is still evident at four and five years of age, and whether children with SLI resemble their typically developing peers in this regard.

In summary, our major goal in the present study was to assess the diagnostic accuracy of two of the Hadley and Short (2005) measures of T/A morpheme use at an older age (four and five years of age) to determine whether they serve as a suitable alternative or supplement to a more conventional composite measure of T/A morpheme use. Our secondary goal was to discover whether the differences in the T/A categories in level of productivity seen at 33 months in typically developing children continue to be seen at four and five years of age in typically developing children and children with SLI.

Method

Participants

A total of 55 spontaneous speech samples from four groups of children provided the data for this study – two groups of children with SLI and two groups who were typically developing (TD). A summary of the ages of each group appears in Table 1. The children in one SLI group and one TD group were younger in age, with ages ranging from 48 to 54 months of age (4;0 to 4;6). The mean age for the younger SLI group ($N = 12$, 6 females) was 51.58 months; for the younger TD children ($N = 15$, 5 females), the mean age was 51.33 months. The children in the remaining SLI and TD groups were older, ranging in age from 60 to 66 months (5;0 to 5;6). The mean age for the older SLI group ($N = 13$, 5 females) was 63.15 months, and the mean age for the older TD group ($N = 15$, 6 females) was 62.60 months. All of the children were from Tippecanoe County, IN, and its adjacent counties. Two participants identified themselves as being of Hispanic ethnicity, two of Asian, and the remaining 51 as White-Caucasian. Based on meeting the criteria described below, these children had participated in one of several previous studies conducted in the Child Language Laboratory at Purdue University (Finneran, Francis, & Leonard, 2009; Krantz & Leonard, 2007; Leonard et al., 2007). Spontaneous speech samples obtained prior to the children’s participation in the experimental tasks of those studies served as the data examined in the present study. All procedures, including the collection, transcription, and analysis of

spontaneous speech samples, had been approved by the Institutional Review Board of the authors' institution.

All participants were monolingual English speakers and passed a bilateral pure tone hearing screening at 500, 1000, 2000, and 4000 Hz at 20 dB. The Columbia Mental Maturity Scale (CMMS; Bergemeister, Blum & Lorge, 1972), a test of nonverbal intelligence, was administered to ensure that all participants exhibited age-appropriate cognitive functioning with scores above 85. Further, all participants passed an oral-mechanism examination according to the Robbins and Klee protocol (1987). Finally, according to parent report, no participants had a history of neurological impairment.

The children with SLI had been previously identified as exhibiting a language impairment. For inclusion in the experimental studies (and in the present study), their scores on the Structured Photographic Expressive Language Test—Second Edition (SPELT—II; Werner & Kresheck, 1983) had to fall below the 10th percentile. In fact, the scores of all of the children with SLI were at or below the 1st percentile. All children in the TD groups scored between the 18th and the 85th percentile on this test.

Spontaneous Speech Samples

Each child was seen individually for a spontaneous speech sample. The samples were obtained while the child and experimenter played with a set of age-appropriate toys. The experimenter made an effort to allow the child to initiate conversational turns by describing the play activities and asking for the child's assistance in selecting particular toys or creating some play scene. However, the experimenter occasionally asked specific questions in the hope that the child would follow the reply with additional, spontaneous utterances. Using this procedure, all children produced at least 152 spontaneous utterances.

Audio-recordings of the children's utterances were transcribed and coded using the Systematic Analysis of Language Transcripts (SALT) software (Miller & Chapman, 2000). The default codes employed by SALT were supplemented to distinguish between copula and auxiliary *is*, *are*, *am*, *was*, and *were* in their contracted and uncontracted forms.

Scoring

Finite verb morphology composite (FVMC)—The FVMC, adapted from Leonard, Miller, and Gerber (1999), was used as the traditional measure of finite verb morphology for comparison with the two measures developed by Hadley and Short (2005). The specific T/A morphemes in the FVMC are copula *is*, *are*, *am*, *was*, *were*, auxiliary *is*, *are*, *am*, *was*, *were*, present third person singular *-s*, and past tense *-ed*. To promote a more appropriate comparison with the Hadley and Short measures, we added three other T/A morphemes to this list, auxiliary *do*, *does*, and *did*. To calculate the FVMC, the total number of appropriately used morphemes from this list was divided by the total number of obligatory contexts for these morphemes, and then this number was multiplied by 100 to arrive at the composite score. One additional rule was added to this calculation procedure; if a child over-regularized a past tense form (e.g., *throwed* instead of *threw*), it was scored as an additional obligatory context and the child was credited with an additional instance of past tense *-ed*. An illustration of this scoring is provided using the abbreviated sample shown in (1).

- (1) Maybe it goes with this zoo? (third person singular *-s*)
 Here's another man that works at the zoo. (copula *is*, third person singular *-s*)
 What does this guy do? (auxiliary *does*)

Is that how it go? (copula *is*, omission of third person singular *-s*)

But what is this for? (copula *is*)\

The popcorn bouncing! (omission of auxiliary *is*)

What are these for? (copula *are*)

I runned fast (over-regularization of past tense *-ed*)

Daddy's a Red Sox fan (copula *is*)

So water's going down (auxiliary *is*)

Within this sample, there are 12 obligatory contexts (including the over-regularized *-ed*). morphemes. Of these 12, this child appropriately produced 10, for a FVMC of 83 (10/12).

Tense Marker Total—Two of the measures outlined by Hadley and Short (2005) were included in the present study because they provided information not readily available from the FVMC and they seemed less vulnerable to ceiling effects for the ages under investigation here. The first was the Tense Marker Total (TMT). This measure assesses 15 different T/A forms: copula *is*, *are*, *am*, *was*, *were*, auxiliary *is*, *are*, *am*, *was*, *were*, auxiliary *do*, *does*, *did*, present third person singular *-s*, and past tense *-ed*. All copula and auxiliary forms were required to be uncontracted or contracted to nominal forms (e.g., *cow's* all gone). Forms contracted to pronominal forms (e.g., *it's* all gone) were excluded. As discussed by Hadley and Short (2005), both correct uses and overregularizations of verb inflections provide evidence of emergence and were therefore included. All other errors in tense marking were excluded. The TMT was the total number of different morphemes out of the 15 that appeared at least once in the sample. Thus, the possible range for the TMT was 0 to 15. Using the same abbreviated sample provided in (1), this child would have a TMT score of 6, with 1 point for copula *is*, 1 for copula *are*, 1 for auxiliary *is*, 1 point for auxiliary *does*, 1 for past tense *-ed*, and 1 for third person singular *-s*.

Productivity Score—The second Hadley and Short (2005) measure was the Productivity Score (PS). This measure is calculated by awarding 1 point for each sufficiently different use of each morpheme, with a maximum of 5 possible points for each of the 5 morpheme categories. For function words, these 5 points could come from 5 productions of the same morpheme with 5 different subjects, or from a combination of different morphemes from the category and multiple uses of the same morpheme with different subjects. These individual morpheme scores are then combined for all 5 morpheme categories, so that the Productivity Scores could range from 0 to 25. For example, using the sample provided in (1), the child would earn 4 points for COPULA BE, for the use of *Is that how it go?*, *But what is this for?*, *What are these for?*, and *Daddy's a Red Sox fan*. Note that *here's* in the utterance *Here's another man that works at the zoo* was not included in the count because the copula *is* contracted with a pronominal form. In addition to the 4 points awarded for the COPULA BE category, the following points would be awarded for the remaining T/A categories reflected in the sample shown in (1): 1 point for AUXILIARY DO, 1 for AUXILIARY BE, 1 for PAST TENSE *-ED*, and 2 for THIRD PERSON SINGULAR *-S*. By adding these values, one arrives at a PS of 9.

Reliability

Approximately 25% of the spontaneous samples, with an even distribution represented from each group, were coded for reliability by a second, independent and blind coder trained to score the FVMC, TMT and PS. The interobserver agreement for each measure was 96.35% for the FVMC, 97.62% for the TMT, and 95.43% for the PS.

Analyses

We examined each of the measures, FVMC, TMT, and PS, first in terms of possible diagnostic category and age differences, and subsequently in terms of the diagnostic accuracy of these measures. To examine diagnostic group and age differences, two-way factorial ANOVAs were used with diagnostic group (TD, SLI) and age (Younger, Older) as between-group variables.

To assess the diagnostic accuracy of the FVMC, TMT, and PS in classifying the participants into TD and SLI groups, logistic regression was used to compute sensitivity and specificity. In addition, positive (LR+) and negative (LR-) likelihood ratios were computed. As described by Dollaghan (2007), the positive likelihood ratio (LR+) “reflects confidence that a positive (disordered or affected) score on a test came from a person who has the disorder rather than from a person who does not” (p. 93). This is calculated by using the formula $LR+ = \text{sensitivity}/(1 - \text{specificity})$. The higher the LR+, the greater the confidence that the score came from a person with the disorder. The negative likelihood ratio (LR-) is then calculated as $LR- = (1 - \text{sensitivity})/\text{specificity}$. This reflects the “confidence that a score in the unaffected range comes from a person who truly does not have the target disorder” (Dollaghan, 2007, p. 93). The lower the LR-, the greater the confidence that the score came from an unaffected person (for a complete description of LRs, see Sackett, Haynes, Guyett, & Tugwell, 1991).

A final analysis was designed to pursue our secondary goal of determining whether differences in the productivity levels of particular T/A categories relative to others would match those observed by Rispoli et al. (in press) for children 33 months of age. The particular productivity measure used by Rispoli et al. (in press) was the previously described Productivity Score, though with each of the five T/A categories taken separately (range = 0 to 5 for each category). A mixed model ANOVA was computed with diagnostic group (TD, SLI) and age (Younger, Older) as between-subjects variables and category type (COPULA BE, AUXILIARY BE, AUXILIARY DO, THIRD PERSON SINGULAR -S, PAST TENSE -ED) as a within-subjects variable. Any interaction was examined through least-significant difference testing.

Results

Finite Verb Morphology Composite

Group differences—The ANOVA for the FVMC revealed a main effect for diagnostic group, $F(1, 51) = 86.02, p < 0.001, \eta_p^2 = 0.63$, with scores significantly lower for the participants with SLI compared to their TD peers. There was no significant difference according to age, $F(1, 51) = 2.84, p = 0.10, \eta_p^2 = 0.09$. Also, the interaction effect for diagnostic group and age was non-significant, $F(1, 51) = 2.13, p = 0.15, \eta_p^2 = 0.14$. Table 2 provides a summary of all means and standard deviations for the two diagnostic groups at each age level.

Diagnostic accuracy—Logistical regression was used to compute the sensitivity and specificity for the FVMC. For the younger age group, both sensitivity (12 of 12 participants) and specificity (15 of 15 participants) were 100%, for a total percentage correct of 100%. Because these values were at 100%, LRs could not be calculated for this group. Table 3 provides a summary.

The FVMC was also highly accurate at predicting the diagnostic category for children in the older age group. Sensitivity was 92.31% (12 of 13 participants) and specificity was 93.33% (14 of 15 participants), for a total percentage correct of 92.86%. The LR+ was 13.85, with a

95% confidence interval (CI) of 2.07 to 92.57, indicating that a score within the affected range was over 13 times more likely to come from a participant with SLI than TD. The LR⁻ was 0.08 with a 95% CI of 0.01 to 0.54, reflecting that a score in the unaffected range was highly unlikely to come from a participant with SLI. The diagnostic accuracy of all measures can be seen in Table 4.

Tense Marker Total

Group differences—For the ANOVA for TMT, a main effect was found for diagnostic group, $F(1, 51) = 31.66, p < 0.001, \eta_p^2 = 0.38$. The TMT scores were significantly lower for the participants with SLI relative to their TD peers. There was no significant difference according to age, $F(1, 51) = 1.41, p = 0.24, \eta_p^2 = 0.03$, nor for the diagnostic group X age interaction, $F(1, 51) = 0.01, p = 0.91, \eta_p^2 = 0.00$. A summary can be seen in Table 2.

Diagnostic accuracy—For the younger age group, sensitivity was 83.33% (10 of 12 participants) and specificity was 86.67% (13 of 15 participants), for a total percentage correct of 85.19%. The LR⁺ was 6.25, with a 95% CI of 1.68 to 23.28, indicating that a score within the affected range was over 6 times more likely to come from a participant with SLI than TD. The LR⁻ was 0.19 with a 95% CI of 0.05 to 0.69, demonstrating the improbability that a score in the unaffected range would come from a participant with SLI (see Table 3).

For the older group, the TMT was slightly less accurate at predicting the diagnostic category. Sensitivity was 76.92% (10 of 13 participants) and specificity was 80.00% (12 of 15 participants), for a total percentage correct of 78.57%. The LR⁺ was 3.86, with a 95% CI of 1.34 to 11.05. The LR⁻ was 0.29 with a 95% CI of 0.10 to 0.80 (see Table 4).

Productivity Score

Group differences—For the ANOVA for PS, a main effect was observed for diagnostic group, $F(1, 51) = 18.05, p < 0.001, \eta_p^2 = 0.40$, indicating that PS scores were significantly lower for the participants with SLI than for their TD peers (see Table 2). There was no significant difference based on age, $F(1, 51) = 3.43, p = 0.07, \eta_p^2 = 0.06$, or for the interaction, $F(1, 51) = 0.23, p = 0.63, \eta_p^2 = 0.00$.

Diagnostic accuracy—For the younger age group, the sensitivity was 66.67% (8 of 12 participants) and specificity was 86.67% (13 of 15 participants), for a total percentage correct of 77.78%. The LR⁺ was 3.4, with a 95% CI of 1.37 to 8.46. The LR⁻ was 0.26 with a 95% CI of 0.07 to 0.93 (see Table 3). Based on these results, the PS alone was not especially accurate in identifying children with SLI in the younger age group (see Table 3).

In contrast, for the older group, the PS adequately predicted the diagnostic category of the children. Sensitivity was 84.62% (11 of 13 participants) and specificity was 80.00% (12 of 15 participants), for a total percentage correct of 82.14%. The LR⁺ was 5.5, with a 95% CI of 1.48 to 20.42, and the LR⁻ was 0.25 with a 95% CI of 0.09 to 0.70 (see Table 4).

Tense Marker Total in Combination with Productivity Score

Because both the TMT and the PS predicted the diagnostic category for either the younger or older age children to an acceptable degree, these measures were combined to determine if predictability would increase in identifying children with SLI and TD at both age groups. Using logistic regression, the TMT + PS combination resulted in a sensitivity of 83.33% (10 of 12 participants) and specificity of 86.67% (13 of 15 participants), for a total percentage correct of 85.19% for the younger age group. The LR⁺ was 6.25, with a 95% CI of 1.68 to 23.28. The LR⁻ was 0.19 with a 95% CI of 0.05 to 0.70. Although producing acceptable

results, the combination of TMT + PS proved no more accurate than the use of TMT alone for these younger children.

For the older group, sensitivity was 84.62% (11 of 13 participants) and specificity was 80.00% (12 of 15 participants), for a total percentage correct of 82.14%. The LR+ was 5.5, with a 95% CI of 1.48 to 20.42. The LR– was 0.25 with a 95% CI of 0.09 to 0.70. The predictability of these two measures combined was roughly the equivalent of the PS alone at categorizing the older group.

Differences Across T/A Categories

Finally, a mixed model ANOVA was pursued to determine whether the differences in relative productivity across the T/A categories resembled those reported by Rispoli et al. (in press) for 33-month-olds. A main effect was found for diagnostic group, $F(1, 51) = 34.01$, $p < 0.001$, $\eta_p^2 = 0.40$, indicating that participants with SLI had lower scores than their TD peers. The main effect for T/A category was also significant, $F(4, 204) = 33.57$, $p < 0.001$, $\eta_p^2 = 0.14$. The main effects for age, $F(1, 51) = 3.43$, $p = .07$, $\eta_p^2 = 0.06$, and the interaction between age and diagnostic group, $F(1, 51) = 0.23$, $p = .63$, $\eta_p^2 = 0.004$, age and T/A category, $F(4, 204) = 0.52$, $p = .72$, $\eta_p^2 = 0.01$, and the three-way interaction of age, diagnostic group and T/A category, $F(4, 204) = 0.85$, $p = .49$, $\eta_p^2 = 0.02$, were all statistically non-significant.

However, we observed a significant interaction between T/A category and diagnostic group, $F(4, 204) = 6.41$, $p = 0.007$, $\eta_p^2 = 0.02$. As can be seen from Figure 1, the two diagnostic groups (collapsed across age level) showed a similar pattern across the T/A categories. Specifically, for the TD group, LSD testing indicated the following pattern of productivity progression: Scores for COPULA BE were significantly higher than for all other categories (p values ranged from .02 to $< .001$, d values range from 0.75 to 1.89). In addition, scores for THIRD PERSON SINGULAR –S were higher than for PAST TENSE –ED ($p < .001$, $d = 1.07$) and for AUXILIARY BE ($p < .001$, $d = 0.91$). Scores for AUXILIARY DO were significantly higher than PAST TENSE –ED ($p < .001$, $d = 1.14$) and for AUXILIARY BE ($p < .001$, $d = 0.98$). However, the latter two categories did not differ ($p = .44$, $d = 0.16$). For the children with SLI, the following pattern of productivity was observed: Scores for COPULA BE were significantly higher than all other morphemes (all p s $< .001$, d s ranged from 1.36 to 2.09). Scores for THIRD PERSON SINGULAR –S were higher than AUXILIARY BE ($p = .02$, $d = 0.56$). There were no significant differences among the scores for AUXILIARY DO, AUXILIARY BE, and PAST TENSE –ED (p s ranged from .14 to .53, d s ranged from 0.15 to 0.39). Clearly contributing to the interaction between T/A category and diagnostic group was the fact that, despite the higher scores of the TD children overall, the two diagnostic groups did not differ in their productivity levels for COPULA BE and PAST TENSE –ED (see Figure 1).

Discussion

All three measures used in the present study – the FVMC, the TMT, and the PS – showed very large differences between the children with SLI and their TD peers, favoring the latter. The more traditional FVMC measure displayed acceptable sensitivity and specificity for both the younger (4;0 to 4;6) and older (5;0 to 5;6) age groups. The TMT also showed acceptable sensitivity and specificity but only for the younger age group; the PS, in contrast, exhibited acceptable levels only for the older age group. When the TMT and PS were combined, acceptable sensitivity and specificity levels were seen for both age groups. However, at each age, the combination failed to increase accuracy levels above those seen for the best measure at that age (TMT for the younger group, PS for the older group). Finally, the differences among T/A categories for both age groups and both diagnostic

groups resembled those reported by Rispoli et al. (in press) for much younger typically developing children. Before turning to the implications of these findings, we discuss some potential limitations of the study.

Some Qualifications

The present study was founded on the assumption that T/A measures from spontaneous speech provide an important perspective on children's grammatical abilities during their daily communication. However, estimates of these abilities can also be obtained through more formal measures. For example, the TEGI (Rice & Wexler, 2001) assesses children's use of many of the T/A morphemes examined in the present study. Only copula and auxiliary *am, was, were*, and auxiliary *did* are not included in the TEGI. If neither spontaneous speech nor a wider range of T/A morphemes is essential, instruments such as the TEGI can be considered as suitable alternatives for the age range studied here.

Another qualification is that the data reported here may be quite specific to a sample size of 152 spontaneous utterances. We found that such a sample size was more than adequate to distinguish the SLI and TD groups at both age 4;0 to 4;6 and age 5;0 to 5;6. However, because both TMT and PS are totals and are thus influenced by each new instance of a scorable T/A morpheme, a much larger sample size might have reduced the differences between the SLI and TD groups. We are confident that the difference between the diagnostic groups would remain even with moderate increases in sample size given that the TD children were not approaching ceiling levels with 152 utterances. However, we cannot rule out the possibility that very large sample sizes would reduce the discriminability of the TMT and PS measures.

An additional qualification is that the sensitivity and specificity values reported here should be viewed as applying to a clinically referred sample. A different standard must be applied when a measure is being used to identify language impairment in a community sample, where the presumed prevalence of SLI is approximately 7% (Tomblin et al., 1997). Of course, for testing large numbers of children in the general community (where most children are expected to be typically developing), language samples such as the ones used here would be an unlikely choice. Nevertheless, it should be kept in mind that these measures show acceptable levels of accuracy when the number of children with SLI in the sample approximates the number of TD children who are assessed.

Yet another qualification pertains to the relationship between the gold standard used to measure sensitivity and specificity (the SPELT-II) and the FVMC, TMT, and PS measures. Although the SPELT-II requires a range of language abilities, children with grammatical deficits are likely to score poorly. Therefore, it might be argued that by employing the SPELT-II as the gold standard, we were selecting children with SLI with significant deficits in grammar, making it more likely that the FVMC, TMT, and PS measures would fare well in sensitivity and specificity. We cannot dismiss this argument. On the other hand, the types of grammatical deficits reflected by poor performance on the FVMC, TMT, and PS measures seem to match a very common, and heritable form of language impairment (Bishop, Adams, & Norbury, 2006).

Finally, one of the advantages of the TMT and PS measures is that they provide greater assurance that estimates of children's T/A morphological abilities are not inflated due to the inclusion of forms that might reflect direct activation rather than grammatical encoding. However, at the ages at which these measures were applied in the present study (4;0 to 4;6 and 5;0 to 5;6), it is possible that forms such as *that's* and *it's* are already parsed as pronominal subject + copula *is* in the child's grammar and hence are no longer unanalyzed wholes. This may well be true, and therefore the exclusion of such forms would represent a

highly conservative approach to evaluating children's T/A morphological abilities. Still, we would argue that the TMT and PS measures would retain another of its major advantages – providing an indication of the diversity of forms and contexts in which children use T/A morphemes. This diversity is simply not provided by the FVMC.

The Contributions of the FVMC, TMT, and PS

The FVMC calculates the percentage of accurate productions out of the total number of obligatory contexts across all 15 T/A morphemes. The TMT measure captures the range of different T/A morphemes produced, and the PS reflects the overall productivity of the five T/A categories that are represented by the 15 individual T/A morphemes. As a collection, these measures seem to provide a very good estimate of a child's ability, as they assess overall accuracy (FVMC), variety (TMT), and productivity (PS).

Our findings for the FVMC essentially replicate earlier investigative efforts by Bedore and Leonard (1998) and Rice (2003). Although our computation of the FVMC was most like that of Bedore and Leonard, those investigators did not include auxiliary *do* forms. Nevertheless, our results were quite similar to theirs. These findings indicate that a measure such as the FVMC can play an important role in identifying children with SLI with grammatical deficits, at least in a clinical referral situation.

It is noteworthy that the TMT exhibited acceptable sensitivity and specificity for the age range of 4;0 to 4;6 and the PS displayed similar positive results for the age range of 5;0 to 5;6. These measures were originally designed for younger ages. The fact that they proved informative in the present study indicates that their foci – T/A morpheme variety and productivity – do not reach asymptote during the preschool years and therefore seem important to monitor throughout this period.

As measures of T/A assessment, the FVMC, TMT, and PS seem to play somewhat different roles. Indeed, an inspection of the data reveals several instances in which two children had identical scores on one of the measures but rather different scores on the remaining measures. For example, two participants in the younger SLI group with identical FVMC scores of 43% attained different TMT scores of 6 and 4, and even more divergent PS scores of 13 and 5. Two other participants from the same group had identical TMT scores of 6 but had FVMC scores of 76% and 40%, and PS scores of 12 and 8. Two additional participants from this group earned identical PS scores of 12, but differed in both FVMC scores (49%, 66%) and TMT scores (5, 7).

One might argue that the FVMC is sufficient for identifying children at risk for language impairment at four and five years of age. However, without close inspection of the samples, it would not be clear which T/A morphemes or categories are in greatest need of clinical attention. The TMT can provide the professional with an indication of the range of T/A morphemes that can be put to use by the child, and the PS can provide a good estimate of the productivity with which the child uses morphemes from each T/A category. Such information seems crucial for planning intervention. Note as well that all three measures can be obtained from the same spontaneous speech sample.

Differences Across T/A Categories

Rispoli et al. (in press) reported differences in younger children's productivity across the T/A categories. COPULA BE forms showed the earliest productivity and AUXILIARY BE appeared to be the last to exhibit productivity. In general, our results are in accord with those of Rispoli et al. These findings can therefore be interpreted as consistent with the assumptions of the Gradual Morphosyntactic Learning Account that shared features of the

morphemes and the types of sentence frames in which the morphemes must be inserted are major contributors to the particular order in which these T/A categories reflect productivity.

Why were differences across T/A categories still apparent at four and five years of age, especially among the TD groups? One likely answer is that the sample size of 152 spontaneous utterances used in the present study was smaller than the sample sizes used by Rispoli et al. This implies, as noted earlier, that if we had obtained larger samples from our participants, some of the differences among the T/A categories might have narrowed. On the other hand, it is clear from a comparison of our data with those of Rispoli et al. that T/A category scores for the TD groups continued to increase from 33 months to four and five years of age, notwithstanding the fact that our sample sizes were smaller. For example, whereas Rispoli et al. reported mean scores of 1.42, 2.74, and 1.89 for AUXILIARY BE, AUXILIARY DO, and THIRD PERSON SINGULAR –S, respectively, our older TD group showed means of 2.73, 4.27, and 4.40, respectively. PAST TENSE –ED showed a less dramatic increase (2.05 reported by Rispoli et al., and 2.80 for our older TD group). We suspect this had to do with the smaller number of obligatory contexts that arose for this T/A category in our sample. Note that this category failed to reveal significant differences between our SLI and TD groups – a surprising finding given that the more traditional measure of percentages of use in obligatory contexts often reveals a significant difference between preschool-aged children with SLI and same-age peers in the use of past tense *-ed* (e.g., Leonard et al., 1992; Rice & Wexler, 1996). (Indeed, the older SLI and TD groups in the present study differed as well, with mean percentages of use in obligatory contexts of 61.54 and 98.46, respectively.) The 152-utterance samples were evidently too small to allow the TD children to accumulate a large enough number of different verbs to be used with the past tense inflection to be distinguishable from the children with SLI in their PAST TENSE –ED category scores. Finally, COPULA BE showed the smallest increase (4.11 reported by Rispoli et al, and 4.73 for our older TD group). Perhaps because COPULA BE is already well established prior to the preschool period, little room is left for subsequent gains in the older age range when 5 is used as the maximum score.

One important finding was the fact that the differences among the five T/A category productivity scores were quite similar for the SLI and TD groups. For both diagnostic groups, the COPULA BE category scores were higher than all other T/A category scores, and the scores for AUXILIARY BE were lower than the scores for at least of two of the remaining T/A categories. Both of these observations are compatible with those of Rispoli et al. (in press). This finding is consistent with the view that children with SLI might get a late start in T/A category use and may continue to show less ability than TD children during the preschool years but may not differ in their profile of development across the categories. Put in terms of the Gradual Morphosyntactic Learning account, despite their lower proficiency, children with SLI are influenced in the same way as TD children by the effects of frequency of occurrence, grammatical feature composition, and type of sentence frame that must accommodate the T/A form. Their manner of transitioning from utterances reflecting direct activation to those involving grammatical encoding seems to be approximately the same as that seen in their TD peers. Such a conclusion appears to be consistent with longitudinal studies using more traditional composite measures of T/A use (Rice et al., 1998) but, in this case, applies to productivity in particular.

Future investigations should continue the study of how productivity changes over time and development, and how these changes may differ in subtle ways between children with SLI and their TD peers. For example, Rispoli, Hadley, and Holt (2009) examined growth in productivity in the spontaneous speech of young TD children using hierarchical linear modeling. Using the entire PS range of 25 by combining the five T/A categories, Rispoli et al. found evidence for instantaneous linear growth at 21 months and an overall acceleration

in productivity through 30 months of age. Although we know from the present study that children with SLI will show lower PS values than their TD peers, and that the same T/A categories are likely to be at the leading edge of development for both diagnostic groups, we do not yet know if the growth curves in productivity for children with SLI will resemble those seen for TD children. Will children with SLI, like their TD counterparts, show a pattern of accelerated growth, interpreted by Rispoli et al. as consistent with a maturational model of acquisition (with a delayed start on the part of the children with SLI)? Or, will growth in productivity for children with SLI appear to be linear, with the lack of acceleration due, perhaps, to less efficient grammatical encoding once this encoding process supplants direct activation as the dominant means of acquiring new grammatical forms?

Conclusions

Although a conventional T/A morphology composite proved to be a good means of distinguishing preschool-aged children with SLI from their typically developing peers, this type of measure is insensitive to important details of the T/A morphology system. Because the diversity (TMT) and productivity (PS) measures introduced by Hadley and Short (2005) showed acceptable diagnostic accuracy for the same age groups and provided more detailed information about T/A morphology, these measures would be appropriate accompaniments to the traditional composite score, and might even serve as a suitable substitute.

These newer measures also provided insight into the developmental progression of T/A productivity. The pattern of least-to-most productive T/A category seen in our preschool TD groups resembled that reported for much younger children by Rispoli et al. (in press), albeit (as expected) with much higher levels of productivity for most categories. In addition, even though the participants with SLI were less productive overall, their T/A productivity showed the same pattern as their TD peers. This finding is important in showing that, despite the more fragile T/A morpheme abilities of children with SLI, these children's grammars, like those of their peers, may be sensitive to the same details that allow some morphemes to develop together (e.g., shared features) and those that cause others to lag slightly behind (e.g., the sentence frame that must accommodate them).

Acknowledgments

The research in this paper was supported in part by Research Grant R01 DC00458 from the National Institute of Deafness and Other Communication Disorders, National Institutes of Health. We thank Patricia Deevy for her assistance with the management of the spontaneous speech samples. We extend our gratitude to Pamela Hadley for answering questions regarding the coding of the Tense Marker Total and Productivity Score. We also thank Kara Bird for her efforts completing the interobserver reliability coding. Finally, we thank the children and families who participated in this study.

References

- Bedore L, Leonard L. Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1185–1192.
- Bishop DVM, Adams C, Norbury CF. Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain & Behavior*. 2006; 5:158–169.
- Bergemeister, B.; Blum, L.; Lorge, I. *Columbia Mental Maturity Scale*. New York: Harcourt Brace Jovanovich; 1972.
- Dollaghan, C. *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Brookes; 2007.
- Finneran D, Francis A, Leonard L. Sustained attention in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2009; 52:915–929.

- Hadley PA, Holt JK. Individual differences in the onset of tense marking: A growth-curve analysis. *Journal of Speech, Language, and Hearing Research*. 2006; 39:984–1000.
- Hadley PA, Short H. The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2005; 48:1344–1362.
- Krantz L, Leonard L. The effect of temporal adverbs on past tense productivity by children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2007; 50:137–148.
- Leonard, L. *Children with specific language impairment*. Cambridge, MA: MIT Press; 1998.
- Leonard L, Bortolini U, Caselli MC, McGregor K, Sabbadini L. Morphological deficits in children with specific language impairment: The status of features in the underlying grammar. *Language Acquisition*. 1992; 2:151–179.
- Leonard L, Deevy P, Kurtz R, Krantz L, Owen A, Polite E, Elam D, Finneran D. Lexical aspect and the use of verb morphology by children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 2007; 50:759–777.
- Leonard LB, Miller C, Gerber E. Grammatical morphology and the lexicon in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 1999; 42:678–689.
- Miller, J.; Chapman, R. *Systematic analysis of language transcripts*. Madison: University of Wisconsin, Language Analysis Laboratory; 2000. (Research Version 6.1a) Computer software
- Plante E, Vance R. Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in the Schools*. 1994; 25:15–24.
- Rice, ML. A unified model of specific and general language delay: Grammatical tense as a clinical marker of unexpected variation. In: Levy, Y.; Schaeffer, J., editors. *Language competence across populations: Toward a definition of specific language impairment*. Mahwah, NJ: Erlbaum; 2003. p. 63-95.
- Rice ML, Wexler K. Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech, Language, and Hearing Research*. 1996; 39:1239–1257.
- Rice, ML.; Wexler, K. *Rice/Wexler Test of Early Grammar Impairment*. San Antonio, TX: Psychological Corporation; 2001.
- Rice ML, Wexler K, Cleave P. Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research*. 1995; 38:850–863. [PubMed: 7474978]
- Rice ML, Wexler K, Hershberger S. Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*. 1998; 41:1412–1431.
- Rispoli M, Hadley PA, Holt JK. The growth of tense productivity. *Journal of Speech, Language, and Hearing Research*. 2009; 52:930–944.
- Rispoli M, Hadley PA, Holt JK. Sequence and System in the Acquisition of Tense and Agreement. *Journal of Speech, Language, and Hearing Research*. (in press).
10.1044/1092-4388(2011/10-0272)
- Sackett, D.; Haynes, R.; Guyatt, G.; Tugwell, P. *Clinical epidemiology*. Boston, MA: Little, Brown; 1991.
- Tomblin JB, Records N, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*. 1997; 40:1245–1260.
- Werner, EO.; Kresheck, JD. *Structured Photographic Expressive Language Test – II*. DeKalb, IL: Janelle Publications; 1983.

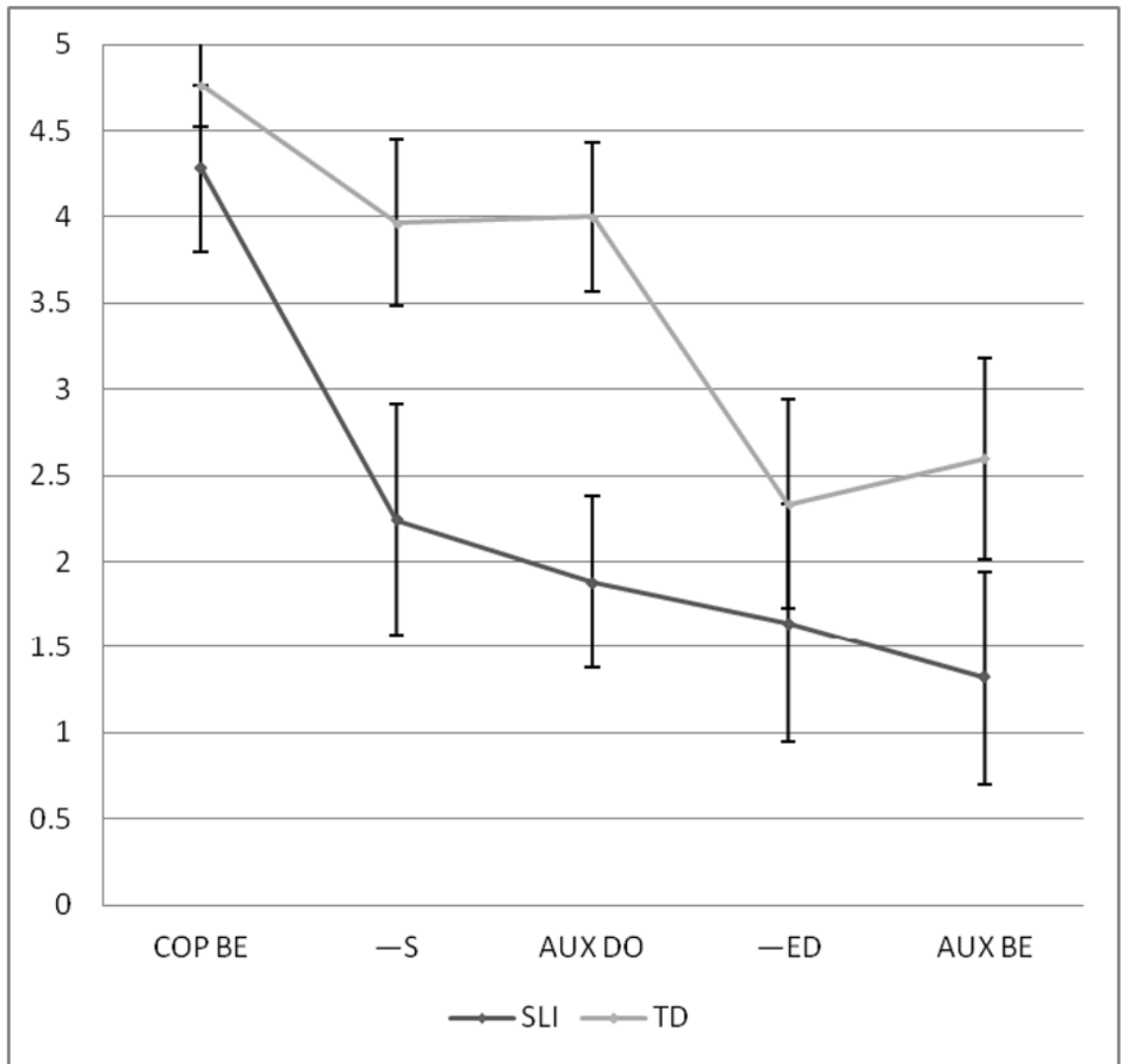


Figure 1. Mean tense/agreement category productivity scores with 95% confidence intervals for the children with SLI and their TD peers, collapsed across younger and older age groups.

Table 1

Means, Standard Deviations, and Ranges for Participant Groups

	N	Mean Age	(SD)	Age Range in Months
<i>Younger</i>				
SLI	12	51.58	2.11	48–54
TD	15	51.33	2.16	48–54
<i>Older</i>				
SLI	13	63.15	1.91	60–66
TD	15	62.60	2.50	60–66

Table 2

Means and standard deviations for the participant groups.

	FVMC		TMT		PS	
	M	SD	M	SD	M	SD
<i>Younger</i>						
SLI (n = 12)	0.56	0.21	5.83	1.75	10.58	4.74
TD (n = 15)	0.96	0.04	8.67	1.59	16.40	3.92
<i>Older</i>						
SLI (n = 13)	0.68	0.19	6.38	2.40	12.08	4.01
TD (n = 15)	0.97	0.05	9.33	1.80	18.93	3.43

Table 3

Diagnostic accuracy for all measures for the younger participant group.

		<u>Clinical Classification</u>				
		SLI	TD	Correct Classification	Likelihood Ratio	
FVMC Classification	SLI	12	0	Sensitivity 100.00%	LR+ N/A	
	TD	0	15	Specificity 100.00%	LR- N/A	
TMT Classification	SLI	10	2	Sensitivity 83.33%	LR+ 6.25	
	TD	2	13	Specificity 86.67%	LR- 0.19	
PS Classification	SLI	8	2	Sensitivity 66.67%	LR+ 3.40	
	TD	4	13	Specificity 86.67%	LR- 0.26	

Table 4

Diagnostic accuracy for all measures for the older participant group.

		<u>Clinical Classification</u>						
		SLI	TD	Correct Classification	Likelihood Ratio			
FVMC Classification	SLI	12	1	Sensitivity	92.31%	LR+	13.85	
	TD	1	14	Specificity	93.33%	LR-	0.08	
TMT Classification	SLI	10	3	Sensitivity	76.92%	LR+	3.86	
	TD	3	12	Specificity	80.00%	LR-	0.29	
PS Classification	SLI	11	3	Sensitivity	84.62%	LR+	5.50	
	TD	2	12	Specificity	80.00%	LR-	0.70	