Identification of Clinical Markers of Specific Language Impairment in Adults

Gerard H. Poll  
The Pennsylvania State University, University Park

Stacy K. Betz  
University of Washington, Seattle

Carol A. Miller  
The Pennsylvania State University, University Park

Purpose: To investigate the usefulness of 3 tasks known to be effective diagnostic clinical markers of specific language impairment (SLI) in children: (a) nonword repetition, (b) sentence repetition, and (c) grammaticality judgments of finiteness marking.

Method: Two groups of young adults, 13 with SLI and 18 with typical language, completed 3 experimental tasks: (a) nonword repetition, (b) sentence repetition, and (c) grammaticality judgments of sentences that were either correct or contained an omitted finiteness marker, an overt agreement error, or an omitted progressive –ing. Analyses included receiver operating characteristic curve analyses and computation of likelihood ratios associated with the use of each task as a clinical marker for SLI, as well as development of a logistic regression model that used multiple tasks as predictors.

Results: Each marker task significantly contributed to classification of adults as affected or unaffected by SLI, with moderate positive and negative likelihood ratios. A combination of the 3 marker tasks was the best predictor of affectedness status with moderate to large likelihood ratios.

Conclusions: The results suggest that SLI persists into adulthood and that effective clinical markers of this disorder are similar to those used to identify SLI in children. Refinement of these tasks to increase their likelihood ratios will improve their usefulness in diagnosing SLI in adults.

KEY WORDS: specific language impairment, adults, assessment procedures

An extensive body of research has focused on specific language impairment (SLI) in children, but few studies have documented the nature of the disorder in adulthood. SLI often goes untreated during the primary and secondary school years (Tomblin et al., 1997; Zhang & Tomblin, 2000), and more than 70% of cases identified in kindergarten persist into young adulthood (Johnson et al., 1999). The persistence of SLI is associated with deficits in academic performance, limited educational attainment, and more extensive unemployment compared with individuals without SLI (Clegg, Hollis, Mawhood, & Rutter, 2005; Records, Tomblin, & Freese, 1992; Young et al., 2002). Zhang and Tomblin (2000) and Johnson and colleagues (1999) have suggested that one reason for the limited rate of intervention for language impairments may be the subtle nature of the deficits: Parents and teachers are less likely to refer a child for suspected language impairment than for a more obvious deficit, such as an articulation disorder. As young adults, individuals with SLI may be even less likely to be identified through observation in informal conversation (Tomblin, Freese, & Records, 1992). Zhang and Tomblin suggested that this issue of underidentification calls for better diagnostic tools.
Several researchers have responded to this need for better assessment tools by exploring clinical markers of language impairment. The term clinical marker was introduced to the language disorders literature by Rice and Wexler (1996) to refer to highly accurate diagnostic tasks that are based on characteristics or behaviors that are indicative of SLI. The notion of markers, clinical markers, or biomarkers existed in the medical literature before the language disorders literature (Bremen, Coffi, Bomba-Ire, Foster, & Herrmann, 1975; Smith, Lipworth, Cree, Spiers, & Winter, 1995). In general, the term marker has been used to denote a sign or symptom that aids diagnostic accuracy by identifying or predicting the presence or absence of disease and that may help define a disease entity (Abeloff, Armitage, Niederhuber, Kastan, & McKenna, 2008; Conti-Ramsden, 2003; Rice, 2003). For example, the presence of a specific protein in the blood has been explored as a marker of pneumonia and has been used to differentiate patients who have pneumonia from those who have simple bronchitis, especially in the face of other conflicting diagnostic data (Smith et al., 1995). Similarly, child language researchers have used the term clinical marker to refer to tasks with high levels of diagnostic accuracy. Stokes, Wong, Fletcher, and Leonard (2006) identified diagnostic accuracy as the lack of score overlap between affected and unaffected groups. Rice (2003) also suggested that ideal clinical markers should have a bimodal distribution. Identification of effective markers for SLI would be invaluable in the diagnostic process used by researchers and clinicians.

In studies of SLI in children, proposed clinical markers have been theoretically based on limitations in cognitive processing mechanisms (e.g., phonological memory) and linguistic knowledge (i.e., finiteness marking; Conti-Ramsden, 2003). Researchers have debated whether these language limitations fall along a continuum with normal performance or are discontinuous from typical performance. It may be that some deficits seen in SLI are continuous with typical performance whereas others are clearly distinct. For example, phonological memory differences may be graded, whereas gaps in linguistic knowledge may result in discontinuous distributions of performance (Joanisse & Seidenberg, 2003). Regardless of the outcomes of this debate, all researchers investigating clinical markers of SLI share the common goal of finding group performance differences that result in diagnostic utility. Performance levels on potential diagnostic measures have historically been measured using a group mean difference approach (e.g., t tests). Only more recently have these analyses focused on quantifying the degree of separation between affected and unaffected individuals (Conti-Ramsden, 2003; Conti-Ramsden, Botting, & Faragher, 2001; Dollaghan & Campbell, 1998; Rice & Wexler, 2001).

### Effective Diagnostic Tools

The medical literature advises diagnosticians to choose tests, signs, or tools for which the results of affected individuals are clearly distinct from the results of unaffected individuals (Sackett, Haynes, Guyatt, & Tugwell, 1991). Therefore, the methods used to determine the usefulness of clinical markers differ from the methods often used in studies of language impairment. The diagnostic accuracy of a new clinical marker is compared to a reference standard, also known as a gold standard. The gold standard is the currently accepted procedure for diagnosing a disorder. Comparison of the new clinical marker and the gold standard can then be assessed using a variety of measures, including sensitivity and specificity, overall classification accuracy, and positive and negative likelihood ratios (LR+ and LR–) (Dollaghan, 2007). Sensitivity measures true positives, or the accuracy of the clinical marker in identifying individuals with a disorder, and specificity measures true negatives, or the accuracy of identifying those without a disorder. Overall classification accuracy is the proportion of true positives and true negatives in the total sample (Conti-Ramsden, 2003).

Another measure that has been recommended for investigating diagnostic methods is the likelihood ratio (Battaglia et al., 2002; Dollaghan, 2007; Dollaghan & Campbell, 1998; Sackett et al., 1991). Likelihood ratios provide advantages over sensitivity and specificity in that they are less affected by differences in base rates of the disorder in the populations of interest because they simultaneously consider the sensitivity and specificity of a test (Dollaghan, 2007). Likelihood ratios can be interpreted in terms of posttest odds, that is, the degree to which a given test result changes the probability that the test-taker is affected with the disorder of interest (McGee, 2002). McGee (2002) provided guidelines for applying likelihood ratios to diagnostic testing, suggesting that an LR+ of 10 increases the posttest probability by approximately 45%, whereas an LR– of 0.1 reduces the posttest probability by 45%. For example, on the basis of previous research, an adult with a family history of language impairment has an estimated pretest probability of being affected with SLI of 30% (Leonard, 1998). Suppose a new clinical marker has an LR+ of 10. The adult’s positive result on the clinical marker test would result in a 75% posttest probability of affectedness. Therefore, LR+ values of 10 or more and LR– values of 0.1 or less are considered extremely informative regarding the presence or absence of a disorder.

To calculate sensitivity, specificity, overall accuracy, and likelihood ratios, a clinician or researcher must first choose an appropriate cutoff score on the clinical marker task that will be used to classify individuals as affected or unaffected. Because a clinical marker aims to identify
two distinct levels of performance (i.e., affected and unaffected), the choice of cutoff score may differ from the score used with normally distributed standardized tests. With normally distributed tests, clinicians typically use a set cutoff score, often 1 or 1.25 SDs below the mean, regardless of the test being administered. Spaulding, Plante, and Farinella (2006) stated that this approach is faulty because the cutoff is arbitrary; it does not consider how accurate that cutoff is in differentiating affected and unaffected individuals. In other areas of health sciences, the notion of diagnostic tests all having normal distributions (and the same cutoff) implies that all diseases have the same prevalence (e.g., 10% of the population is always affected), a conclusion that is not supported by the facts (Sackett et al., 1991).

When using clinical markers, one can determine a cutoff score by observing the receiver operating characteristic (ROC) curve, a graphical means of determining the diagnostic power of a test (Sackett et al., 1991). The ROC curve plots pairs of sensitivity rates against 1 – specificity rates for each cutoff score. Cutoff scores for a clinical marker task determined by ROC curves do not have the inherent arbitrariness that set cutoff points such as −1 SD have.

**Clinical Markers of SLI in Children**

Three types of clinical markers have been explored in children with SLI: (a) nonword repetition, (b) sentence repetition, and (c) grammaticality judgments of finiteness marking. Nonword repetition tasks, which tap both short-term phonological memory and phonological processing, have been shown to result in a much higher likelihood of correct identification of children with SLI when compared with a traditional standardized test. Using a cutoff of fewer than 70% of phonemes correctly imitated, Dollaghan and Campbell (1998) found an LR+ of 25.15 for children ages 5 through 12 years. Using the same nonword repetition task, Ellis Weismer and colleagues (2000) also found clear deficits in performance for children with SLI, ages 7 through 8 years, but the LR+ was only 2.78. High values of sensitivity and specificity for nonword repetition performance have also been found using the Children’s Test of Nonword Repetition (Gathercole & Baddeley, 1996) with 5- and 11-year-old children (Conti-Ramsden, 2003; Conti-Ramsden et al., 2001). Despite the discrepancy in the degree of classification accuracy among previous studies, nonword repetition performance remains a potential clinical marker for SLI.

Sentence repetition tasks have been identified as clinical markers of SLI in children by multiple researchers (Conti-Ramsden et al., 2001; Stokes et al., 2006), and they show high levels of sensitivity and specificity. Unlike nonword repetition and grammaticality judgments of finiteness marking, sentence repetition does not align clearly with a theoretical perspective on the mechanisms underpinning SLI. Sentence repetition may tap short-term memory and information-processing abilities, as well as grammatical knowledge (Betz, Rice, Tomblin, & Chen, 2002; Tomblin et al., 1992). When sentences exceed the capacity for rote serial recall, they are thought to be reconstructive, or reflective of the grammatical system of the individual repeating the sentence (Bley-Vroman & Chaudron, 1994). Conti-Ramsden et al. (2001) found that, when using a cutoff score of the 16th percentile, sentence repetition was the most effective marker task, with a sensitivity of 90% and specificity of 85%, in their study of children with SLI at age 11.

The third potential clinical marker of SLI is production and grammaticality judgments of finiteness marking. Previous studies have shown that children with SLI are more likely than typically developing peers to omit obligatory finiteness markers (Rice & Wexler, 1996; Rice, Wexler, & Cleave, 1995; Rice, Wexler, & Redmond, 1999). These same studies have found that affected and unaffected children were unlikely to use overt incorrect finiteness markers; that is, children will omit these markers, but when they do use them will do so correctly. Rice, Wexler, and Redmond (1999) found parallel results for judgments of finiteness marking. Children with SLI often accepted omitted finiteness markers, as in “*He eat toast,” but they were less likely to accept incorrect forms of finiteness marking such as “*I likes toast.” Furthermore, children with SLI were likely to detect errors involving an omitted progressive –ing, as in “*He is cough,” which is not a finiteness marker. Rice and Wexler (2001) developed a standardized test to diagnose language impairment on the basis of deficits in production and grammaticality judgments of finiteness markers. The results from the standardization sample indicated high levels of sensitivity and specificity, near or above 80%, across the age range of 3 through 8 years. Conti-Ramsden et al. (2001) and Conti-Ramsden (2003) have also found sensitivity levels of 63% through 74% and specificity values of 89% through 91% for tasks involving production of past-tense and third person singular present-tense morphemes. These results suggest that all three tasks—nonword repetition, sentence imitation, and grammaticality judgments and production of finite verb morphology—are promising clinical markers of SLI in children.

Conti-Ramsden and colleagues (Conti-Ramsden, 2003; Conti-Ramsden et al., 2001) have explored whether combinations of marker tasks may improve diagnostic accuracy. Both studies found modest gains by using task combinations. For example, using a criterion of a low score on either sentence repetition or nonword repetition, sensitivity was 96%, and specificity was 78%. In her 2003 study, Conti-Ramsden reported that a discriminant
model including both nonword repetition and a past-tense marking task had better classification accuracy (sensitivity of 81%, specificity of 91%) than did nonword repetition alone. Therefore, the best diagnostic tool may be a combination of clinical marker tasks.

**Diagnosis of SLI in Adulthood**

Similar to SLI in children, there are no generally accepted criteria for assessment of SLI in adults. Only a few studies have assessed and characterized deficits in adults with SLI (Clegg et al., 2005; Johnson et al., 1999; Plante, Shenkman, & Clark, 1996; Tomblin et al., 1992). Most of these studies have diagnosed adults on the basis of a positive history of SLI in childhood and a battery of behavioral assessments. Tomblin and colleagues (1992) identified a battery of tests and scores that best discriminated participants with a history of SLI from those in a control group with no positive history; however, no objective data concerning the prior language abilities of the control group were analyzed. The testing batteries that best differentiated the groups were an in-person set of measures, including the Peabody Picture Vocabulary Test—Revised (PPVT–R; Dunn & Dunn, 1981), speaking rate, the Modified Token Test (de Renzi & Faglioni, 1978; Morice & McNicol, 1985), and written spelling, as well as a set of measures administered via the phone that included sentence repetition, word association, speaking rate, and oral spelling.

Johnson et al.’s (1999) study was based on a large, population-based sample of children followed to adulthood. At both the kindergarten and young adult times of measure, these investigators classified individuals as having language impairment on the basis of a criterion of 1 SD or more below the mean on at least one of two standardized language measures: either the PPVT–R or the Test of Adolescent and Adult Language—3 (TOAL–3; Hammill, Brown, Larsen, & Wiederholt, 1994). The study also included a clinical classification by speech-language pathologists, who tended to use more stringent criteria to classify individuals as impaired. The subset of participants classified by the speech-language pathologists as impaired was not significantly different from the other investigator-classified participants in terms of rate of normal language functioning in young adulthood or levels of academic achievement. Similarly, Plante et al. (1996) found that self-reported history of SLI underestimated the presence of SLI, as assessed by Tomblin et al.’s (1992) battery.

Sackett and colleagues (1991) suggested that, in cases where there is no definitive diagnostic criteria, another way to arrive at a gold standard for the disorder is to consider the consequences that logically follow from the disorder. In the case of SLI in adults, the logical consequences would be functional impacts, such as poor school performance, linked to weak language ability. The academic outcomes for participants in Johnson et al.’s (1999) study were evaluated at age 19. The individuals with language impairment had “clear and pervasive differences” in academic outcomes, including a higher risk for learning disabilities (Young et al., 2002), compared with the typical-language (TL) group. Given that Johnson et al.’s diagnostic criteria align with the logical consequences of language impairment, and that previous studies have found that many individuals with SLI are not diagnosed with language impairments (Tomblin et al., 1997), the gold standard for the present study was the diagnostic criteria Johnson et al. used, without a requirement for a positive history of language impairment. These participant selection criteria will also strengthen inferences about the potential effectiveness of clinical markers with this population: If marker tasks distinguish a mildly impaired group from a TL group, they are more likely to be applicable to individuals with more severe impairment as well. Furthermore, it may prove beneficial to identify impairments on the mild end of the severity scale if those impairments result in functional limitations, as Johnson et al.’s work suggested.

Overall, the studies just discussed have found broad language deficits in adults with SLI, including phonological processing, grammar, and semantics. Adults with SLI performed poorly on measures of receptive and expressive grammar (Johnson et al., 1999; Tomblin et al., 1992) and receptive and expressive vocabulary (Clegg et al., 2005; Tomblin et al., 1992). Deficits in phonological processing were also found, including poor performance on nonsense word repetition and pig latin tasks (Clegg et al., 2005; Young et al., 2002). Although small, the literature base for understanding SLI clearly suggests that the disorder continues into adulthood, and additional diagnostic methods are needed.

**Adult Performance on Clinical Marker Tasks**

Given the need for additional and more precise tools for the diagnosis of SLI in adulthood, in this study we explored whether clinical marker tasks shown to be effective for children with SLI would also be effective in diagnosing SLI in adults. Dollaghan and Campbell’s (1998) nonword repetition task has been shown to clearly distinguish children with SLI from their TL peers (Graf Estes, Evans, & Else-Quest, 2007). The performance of adults with SLI has not been documented for this task; however, using the Nonword Memory Test (Gathercole & Baddeley, 1996), Clegg et al. (2005) found differences between adults with TL and adults with a history of severe language impairment as a child. Barry, Yasin, and Bishop (2007) found that a two- to five-syllable nonword repetition task from the NEPSY (Korkman, Kirk, & Kemp,
1998) classified parents of children with language and literacy impairments with 70% sensitivity and 79% specificity. The contribution of verbal memory deficits to poor nonword repetition performance has not been well defined. A study conducted by Gray (2003) failed to find significant correlations between nonword repetition and verbal memory span. This suggests that nonword repetition may involve capacities beyond short-term memory storage in children. Investigating verbal memory in adults would provide evidence on the contributions of short-term and working memory to nonword repetition performance.

Sentence-imitation tasks have been found to differentiate children with and without SLI (Betz et al., 2002; Conti-Ramsden et al., 2001). Although sentence imitation has been less studied in adults with SLI, these tasks do occur on standardized tests such as the TOAL–3, which are used to diagnose language impairments in adults. In addition, Tomblin et al. (1992) found differences between adults with and without a history of language impairment on the sentence-imitation measure from the Multilingual Aphasia Examination (Benton & Hamsher, 1978). These results suggest that sentence-imitation performance may be a clinical marker of SLI in adults.

Finiteness marking, the third proposed clinical marker of childhood SLI, has not been investigated in adults with SLI. For children, a grammaticality judgment task of finiteness marking distinguished children with SLI age 6–8 from their TL peers (Rice et al., 1999). It is not clear whether deficits in knowledge of tense and agreement marking are resolved in adulthood (Leonard, 1998). Deficits with finiteness marking may continue in adulthood for individuals with SLI but would become apparent only under more demanding conditions (Leonard, Miller, & Gerber, 1999). One way a grammaticality task could be made more demanding is to develop more complex stimuli using relative clauses modifying the subject. Subject–object relative clauses are a later developing form in the grammars of children and have been shown to be challenging to comprehend for older children and for adults with aphasia (Caplan, Baker, & DeHaut, 1985; O’Grady, 1997; Scott, 1988). Therefore, we investigated grammaticality judgments of this type as potential clinical markers of adult SLI.

In summary, the primary questions for this study are the following: First, do Dollaghan and Campbell’s (1998) nonword repetition task, the Recalling Sentences task of the Clinical Evaluation of Language Fundamentals, Third Edition (CELF–3; Semel, Wiig, & Secord, 1994), and a grammaticality judgment task focused on omitted finiteness distinguish adults with SLI from their unaffected peers? Second, which of the clinical marker tasks or combination of tasks best differentiate adults with SLI from TL peers?

### Method

#### Participants

Thirty-one adult native English speakers between the ages of 18;0 (years;months) and 25;11 participated in this study. All participants were current or recent students at a vocational postsecondary school in central Pennsylvania. A vocational school setting was selected based on evidence that the student population was more likely to include a greater proportion of students with learning disabilities than a 4-year university (Horn & Berktold, 1999). Written informed consent was obtained from all participants before testing.

All participants obtained a standard score of 80 or higher on a short-form performance intelligence quotient (PIQ) of the Wechsler Adult Intelligence Scale—III (WAIS–III; Wechsler, 1997). The PIQ was estimated from the Picture Completion, Block Design, and Digit Symbol Coding subtests using the approach outlined by Ryan and Ward (1999; Ward, 1990). The Digit Span subtest of the WAIS–III, including both digits forward and backward, was administered to measure verbal short-term and working memory, but the Digit Span scores did not affect participant classification. All participants also passed a hearing screening in at least one ear at 25 dB HL at 500, 1000, 2000, and 4000 Hz.

Adults were classified into two groups: (a) a group with SLI and (b) a group with TL abilities. Classification of participants followed the procedures of Johnson et al. (1999). We used the following two criteria to classify participants in the SLI group: (a) a score of 1 SD or more below the mean on the spoken language quotient (SLQ) of the TOAL–3 or 2 SDs or more below the mean for any of the four subtests that make up the SLQ, or (b) a score of 1 SD or more below the mean on the PPVT–R. Eleven of 13 members of the SLI group qualified on the basis of an SLQ of 85 or below, 1 qualified on the basis of a PPVT–R standard score below 85, and 1 qualified on the basis of a TOAL–3 subtest standard score more than 2 SDs below the mean.

Participants who scored within or above 1 SD from the mean on the TOAL–3 SLQ and PPVT–R, and who scored within or above 2 SDs from the mean on all subtests of the TOAL–3, were classified as members of the TL group. Participants in the TL group also had no self-reported history of language therapy, special education, significant academic difficulty, or a known diagnosis of a language or cognitive disorder. A positive history of language impairment was not required for inclusion in the SLI group because Tomblin and colleagues (1997) found that fewer than one third of children with SLI had previously been diagnosed. One member of the SLI group reported a history of late talking and language intervention, and one reported previous articulation...
therapy. Eight members of the SLI group reported a history of academic difficulties, primarily with reading. Six of the TL group members reported a history of speech-language intervention; however, they had received services for either articulation or fluency disorders, not language.

Summary statistics for the two participant groups are shown in Table 1. The TL and SLI groups did not differ in terms of age, \(t(29) = 0.484, p = .632\) (two-tailed), or PIQ, \(t(29) = 0.673, p = .507\) (two-tailed). One member of the TL group identified herself as Hispanic, whereas all other participants described their race as White. The SLI group consisted of nearly equal numbers of men and women.

### Experimental Tasks

Participants completed three experimental tasks corresponding to the three clinical marker tasks of interest. The first task was nonword repetition. The stimuli, administration, and scoring followed the process outlined by Dollaghan and Campbell (1998). The 16 stimuli consisted of equal numbers of one-, two-, three-, and four-syllable nonwords. The stimuli were audio-recorded and presented via headphones using a Marantz PMD 201 cassette recorder. All participants heard each nonword one time. Responses were audio-recorded using a Marantz PMD 660 digital recorder and later transcribed using broad phonetic transcription. All transcriptions were completed first by one researcher and then reviewed by a second researcher to arrive at a final consensus transcription. The dependent variables were the total percentage of phonemes correct and percentage correct for each of the four syllable lengths.

The second experimental task, sentence repetition, was taken from the Recalling Sentences subtest of the CELF–3. Administration began with Item 7, as directed by the test guidelines, items were repeated. Training items were sentences, 50 experimental sentences, and 20 filler sentences with varying syntactic structures included but not analyzed. The vocabulary for all sentences consisted of words with a frequency count greater than 10 based on Kuščer and Francis (1967). The order of item presentation was chosen to vary sentence types and avoid rote responses from participants. Each type of experimental sentence was intentionally constrained to a fixed length in syllables to avoid variation based purely on memory constraints. More complex sentences with finiteness omissions and errors were included because responses to these items were expected to better reveal differences in adult participants than simple sentences with the same errors. Fewer complex missing progressive sentences were included because these were control sentences, designed to provide evidence that all participants understood the task.

The 50 experimental items included the following seven components: (a) 5 simple sentences containing a finiteness omission, (b) 5 simple sentences with correct finiteness marking, (c) 10 complex sentences containing a finiteness omission, (d) 10 complex sentences containing an ungrammatical finiteness marker, (e) 10 complex sentences with correct finiteness marking, (f) 5 complex sentences with a missing progressive –ing, and (g) 5 complex sentences with correct use of progressive –ing. All complex sentences contained a subject–object relative clause. The grammatical errors were always present in the main clause, not the relative clause. An additional 20 filler sentences with varying syntactic structures were included but not analyzed. The vocabulary for all sentences consisted of words with a frequency count greater than 10 based on Kuščer and Francis (1967). The order of item presentation was chosen to vary sentence types and avoid rote responses from participants. Each type of experimental sentence was intentionally constrained to a fixed length in syllables to avoid variation based purely on memory constraints. More complex sentences with finiteness omissions and errors were included because responses to these items were expected to better reveal differences in adult participants than simple sentences with the same errors. Fewer complex missing progressive sentences were included because these were control sentences, designed to provide evidence that all participants understood the task. The stimuli given are presented in the Appendix.

All tasks were administered individually in a quiet room at the participants’ school. Order of task presentation was varied across participants. Two sessions, one at the beginning of data collection and one at the end, were video recorded and reviewed by a second researcher to ensure consistency of administration. The results of all participant selection tests as well as the nonword

### Table 1. Summary of subject selection measures.

<table>
<thead>
<tr>
<th>Measure</th>
<th>TL⁹</th>
<th>SD</th>
<th>SLIᵇ</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>21</td>
<td>1.77</td>
<td>21</td>
<td>1.82</td>
</tr>
<tr>
<td>PIQ</td>
<td>103</td>
<td>12.32</td>
<td>101</td>
<td>11.78</td>
</tr>
<tr>
<td>TOAL–3 SLQ</td>
<td>103</td>
<td>5.82</td>
<td>78</td>
<td>8.51</td>
</tr>
<tr>
<td>PPVT–R</td>
<td>103</td>
<td>9.41</td>
<td>98</td>
<td>9.14</td>
</tr>
</tbody>
</table>

Note. TL = typical language; SLI = specific language impairment; PIQ = performance intelligence quotient; TOAL–3 SLQ = Test of Adolescent and Adult Language—3 spoken language quotient; PPVT–R = Peabody Picture Vocabulary Test—Revised.

⁹N = 18 (9 women, 9 men). ᵇN = 13 (7 women, 6 men).
repetition and sentence repetition tasks were double scored by a second researcher to ensure scoring reliability. Any scoring differences were reviewed by both researchers until an agreement was reached.

**Analysis**

Measures of diagnostic accuracy were computed for each of the marker tasks. The cutoff for a positive result was identified by examination of an ROC curve. The cutoff that maximized classification accuracy, where sensitivity plus specificity divided by 2 is largest, was selected (Sackett et al., 1991). With reference to Figure 1, sensitivity was computed as the proportion of individuals in the SLI group with a positive clinical marker result (A/A + C)). Specificity was computed as the proportion of individuals in the TL group with a negative clinical marker result (D/[B + D]; Dollaghan, 2007). Overall classification accuracy was computed as the sum of true positives and true negatives divided by the total sample ([A + D]/[A + B + C + D]; Conti-Ramsden, 2003). Likelihood ratios were calculated from the sensitivity and specificity findings. The LR+ was calculated as the ratio of true positives to false positives (sensitivity/[1 – specificity]). The LR– was calculated as the ratio of false negatives to true negatives ([1 – sensitivity]/specificity).

**Results**

We investigated whether adults with SLI could be distinguished from unaffected adults using tasks shown to be effective in identifying children with SLI. First, would nonword repetition, sentence repetition, and grammaticality judgments of finiteness markers as individual tasks accurately classify affected and unaffected adults? A second question was whether a combination of these clinical marker tasks would more accurately classify the language status of the participants than individual tasks. To address this second question, we performed a sequential logistic regression. We then computed classification accuracy metrics for the final model.

**Nonword Repetition**

Adults with SLI were expected to be distinguished from their TL peers using the nonword repetition task. The group means for total percentages of phonemes correct were SLI = 91.1% (SD = 3.52) and TL = 93.9% (SD = 3.17). No individual from either group scored below 85%. At one- and two-syllable nonword lengths, both groups performed near ceiling levels of 99% correct. Group performance was more distinct for three- and four-syllable nonwords. The means for three-syllable nonwords were SLI = 95.6% (SD = 2.59) and TL = 98.4% (SD = 1.83); for four-syllable nonwords the means were SLI = 80.1% (SD = 8.17) and TL = 85.8% (SD = 7.19).

Given indications of group performance differences for total percentage of phonemes correct as well as for performance at the three- and four-syllable nonword levels, we performed an ROC curve analysis to determine the optimal cutoff points for these measures. Classification accuracy measures are summarized in Table 2. The total phonemes and three-syllable measures resulted in better overall classification accuracy than did the four-syllable measure. At a cutoff of 92% of phonemes correct, total phonemes correct resulted in the best specificity and a moderate LR+ using Dollaghan’s (2007) guidelines, whereas at 98% of phonemes correct, the three-syllable measure resulted in a moderate LR–.

The relationships between nonword repetition and the forward and backward digit span raw scores were analyzed to better understand the contributions of short-term and working memory to nonword repetition performance. On the digits forward task, group means were as follows: SLI = 9.1 (SD = 1.9) and TL = 10.6 (SD = 2.0), SLI < TL, t(29) = 2.07, p = .05 (two-tailed), d = 0.76. On the digits backward task, group differences were more pronounced: SLI = 5.3 (SD = 1.3) and TL = 7.0 (SD = 2.2), SLI < TL, t(29) = 2.47, p = .02 (two-tailed), d = 0.94. For the SLI group, neither the digits forward task (r = .19, p = .55) nor the digits backward task (r = -.33, p = .27) was significantly correlated with the total percentage of phonemes correct. For the TL group, the digits forward task was not significantly correlated with nonword repetition performance (r = .30, p = .22), but the digits backward score was significantly and positively correlated (r = .48, p = .05). Neither the digit span tasks nor the nonword repetition task had floor or ceiling effects that materially affected correlations.

**Sentence Repetition**

In regard to the sentence repetition task raw scores, neither group performed near the ceiling level of 78. The
group means were 57.9 (SD = 7.40) for the SLI group and 69.1 (SD = 5.82) for the TL group. Examination of the ROC curve indicated that a cutoff score of 62.5 would maximize overall classification accuracy, which at 87% was superior to that of the nonword repetition measures (see Table 2). Both sensitivity and specificity exceeded 80%, and the LR+ and LR– were well into the moderate range based on Dollaghan’s (2007) standards for classification accuracy.

We examined the correlations between digit span measures and CELF–3 Recalling Sentences raw scores. Both the direction and magnitude of the correlations across groups were similar. For the SLI group, neither the digits forward ($r = .52, p = .07$) nor the digits backward measures ($r = .36, p = .23$) reached statistical significance. For the TL group, both the digits forward ($r = .50, p = .03$) and digits backward ($r = .47, p = .05$) results were significantly correlated to the sentence repetition scores.

The TOAL–3 SLQ was the primary determinant of participant classification as a member of the SLI group. The SLQ was based on four subtests, including a sentence repetition task. For the SLI group, the correlation between the TOAL–3 sentence repetition subtest and the CELF–3 Recalling Sentences task ($r = .69, p = .01$) was significant. Therefore, poor performance on the CELF–3 Recalling Sentences by the SLI group may have been due to a sentence imitation task being a factor in initial group classification. To better understand whether significant group differences on the CELF–3 Recalling Sentences task were influenced by the use of the TOAL–3 sentence repetition subtest as a classification measure, we reclassified the SLI participants on the basis of their scores on the three remaining subtests making up the TOAL–3 SLQ. Participants were placed in one group if the average of their scores on the three remaining subtests was 1 SD or more below the mean or if any one remaining subtest score was 2 SDs or more below the mean. Five participants met these criteria. The remaining 8 participants, whose scores did not meet these criteria, were placed in a second group. The scores of the two resulting small groups did not meet the distributional assumptions for parametric testing; therefore, we conducted a Mann–Whitney U test to determine whether the groups differed in their performance on the CELF–3 Recalling Sentences. The test revealed no significant difference between these new groups on the CELF–3 Recalling Sentences ($U = 15, z = -0.81, p = .42$). In addition, the TL group ($Mdn = 71, M = 69$) performed better than both the SLI group that met the new criterion of poor performance on three subtests of the TOAL–3 SLQ ($Mdn = 61, M = 60, U = 16, z = -2.2, p = .03$) and the SLI group that did not meet this criterion ($Mdn = 59, M = 57, U = 8, z = -3.6, p < .01$).

### Grammaticality Judgment

For the grammaticality judgment task, there were four outcome variables, all expressed as $A'$ statistics, computed for each sentence type using the procedure outlined by Grier (1971). The $A'$ measure transformed the participants’ “good” and “not so good” judgments into a measure equivalent to percentage correct for an unbiased forced-choice task. The $A'$ statistic has a range from .50 to 1.00, where .50 is equivalent to random chance and 1.00 is perfect discrimination of well-formed sentences.

#### Table 2. Sensitivity, specificity, positive likelihood ratios, negative likelihood ratios, and classification accuracy for differentiating SLI and TL adults for each potential clinical marker task.

<table>
<thead>
<tr>
<th>Task</th>
<th>Cutoff$^b$</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>LR+</th>
<th>LR–</th>
<th>Overall accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonword repetition$^b$</td>
<td>Total PPC</td>
<td>92</td>
<td>.615</td>
<td>.778</td>
<td>2.77</td>
<td>.495</td>
</tr>
<tr>
<td></td>
<td>Three-syllable PPC</td>
<td>98</td>
<td>.846</td>
<td>.556</td>
<td>1.91</td>
<td>.277</td>
</tr>
<tr>
<td></td>
<td>Four-syllable PPC</td>
<td>84</td>
<td>.692</td>
<td>.667</td>
<td>2.08</td>
<td>.462</td>
</tr>
<tr>
<td>Sentence repetition$^c$</td>
<td></td>
<td>62.5</td>
<td>.846</td>
<td>.889</td>
<td>7.62</td>
<td>.173</td>
</tr>
<tr>
<td>Grammaticality judgments$^d$</td>
<td>Simple omitted finiteness</td>
<td>.90</td>
<td>.231</td>
<td>1.0</td>
<td>$\infty$</td>
<td>.769</td>
</tr>
<tr>
<td></td>
<td>Complex omitted finiteness</td>
<td>.95</td>
<td>.538</td>
<td>.944</td>
<td>9.607</td>
<td>.489</td>
</tr>
<tr>
<td></td>
<td>Complex bad agreement</td>
<td>.95</td>
<td>.462</td>
<td>.833</td>
<td>2.677</td>
<td>.646</td>
</tr>
<tr>
<td></td>
<td>Complex missing progressive</td>
<td>.90</td>
<td>.077</td>
<td>1.0</td>
<td>$\infty$</td>
<td>.923</td>
</tr>
</tbody>
</table>

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

versus sentences with errors. Because the score distributions were nonnormal, we conducted nonparametric tests; the results are reported in Table 3. Group differences were judged significant at the .01 level to adjust for family-wise error rate.

Only one member of the TL group did not perfectly discriminate the complex omitted finiteness sentences, whereas a majority of SLI group members made judgment errors in this condition. As expected, this condition resulted in a significant group difference and a large effect size ($r = .54$; Cohen, 1992). The remaining sentence types were not expected to distinguish affected from unaffected individuals but instead would indicate whether the pattern of results paralleled that of young children. Although not statistically significant, the simple omitted finiteness and complex bad-agreement sentences resulted in medium effect size differences between the groups. For the complex missing-progressive sentences, all members of the TL group achieved perfect discrimination and only one member of the SLI group made a judgment error.

ROC curve analysis indicated that performance on the complex omitted finiteness sentences resulted in the best classification accuracy of the four sentence types (see Table 2). This was the result of a very high degree of specificity for the task, which contributed to an extremely positive LR+ based on Dollaghan’s (2007) standards. The remaining sentence types, as expected, resulted in poorer classification accuracy.

**Predictive Value of Combinations of Tasks**

Because each of the three types of diagnostic tasks explored in this study contributed to classification accuracy for adults with SLI, the next consideration was whether a combination of these tasks would result in better diagnostic accuracy than any one task individually. To evaluate this possibility, the most discriminating measure from each of the marker tasks was entered into a sequential logistic regression. We evaluated the set of predictors that yielded the best model using Akaike’s information criterion (AIC; Kutner, Nachtsheim, Neter, & Li, 2005), with a lower AIC indicating a better model.

Because the sentence repetition task had the best overall classification accuracy, we entered the standardized raw score of the sentence repetition task as the first predictor. Standardization of predictors enhanced the interpretability of cross-task comparisons. A summary of the model-building steps is presented in Table 4. The one-variable model was significant, so we entered the standardized $A'$ statistic from the grammaticality judgment task using the complex omitted finiteness sentences. The likelihood ratio test, which indicates the significance of the new variable given the presence of the other predictors, suggested that this variable also enhanced classification accuracy. The addition of the nonword repetition measure, as measured by total percentage of phonemes correct, did not significantly contribute to the model and increased the AIC, so we dropped it and instead entered the standardized three-syllable percentage of phonemes correct measure. This predictor significantly contributed to classification accuracy and reduced the AIC further. The final model included standardized predictors for sentence repetition, the complex omitted finiteness sentence grammaticality judgment, and the three-syllable nonword repetition measure.

### Table 3. Medians and Mann–Whitney U tests of group differences between the TL and SLI participants for the grammaticality judgment (GJ) tasks.

<table>
<thead>
<tr>
<th>Variable</th>
<th>TL Mn</th>
<th>SLI Mn</th>
<th>U</th>
<th>Z</th>
<th>$p^a$</th>
<th>$r^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GJ simple omitted finiteness $A'$</td>
<td>1.00</td>
<td>1.00</td>
<td>90</td>
<td>-2.1</td>
<td>.04</td>
<td>.38</td>
</tr>
<tr>
<td>GJ complex omitted finiteness $A'$</td>
<td>1.00</td>
<td>.95</td>
<td>59</td>
<td>-3.0</td>
<td>.00</td>
<td>.54</td>
</tr>
<tr>
<td>GJ complex bad agreement $A'$</td>
<td>1.00</td>
<td>1.00</td>
<td>78</td>
<td>-2.0</td>
<td>.05</td>
<td>.35</td>
</tr>
<tr>
<td>GJ complex missing progressive $A'$</td>
<td>1.00</td>
<td>1.00</td>
<td>108</td>
<td>-1.2</td>
<td>.24</td>
<td>.21</td>
</tr>
</tbody>
</table>

*a*Two-tailed. *b*Effect size.

### Table 4. Sequential logistic regression predicting probability of being affected with SLI.

<table>
<thead>
<tr>
<th>Variable entered</th>
<th>AIC</th>
<th>LR $\chi^2$</th>
<th>df</th>
<th>$p$</th>
<th>$R^2$</th>
<th>n*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence repetition</td>
<td>27.171</td>
<td>16.9</td>
<td>1</td>
<td>.000</td>
<td>56%</td>
<td></td>
</tr>
<tr>
<td>Complex omitted finiteness $b$</td>
<td>25.614</td>
<td>4.6</td>
<td>1</td>
<td>.032</td>
<td>65%</td>
<td></td>
</tr>
<tr>
<td>Nonword repetition, total PPC</td>
<td>26.296</td>
<td>2.3</td>
<td>1</td>
<td>.129</td>
<td>68%</td>
<td></td>
</tr>
<tr>
<td>Total PPC removed, entered</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonword repetition, three-syllable PPC</td>
<td>22.683</td>
<td>4.9</td>
<td>1</td>
<td>.027</td>
<td>75%</td>
<td></td>
</tr>
</tbody>
</table>

Note. Akaike’s information criterion (AIC) and $R^2$ are for models with all variables present on that step; LR tests are for the contribution of the variable added on that step given the variables already in the model. The final model included the z scores for the sentence repetition, grammaticality judgments of complex omitted-finiteness sentences, and three-syllable nonword repetition PPC.

*a*Nagelkerke’s $R^2$. *b*Complex omitted finiteness grammaticality judgment sentences, z-scored $A'$ statistic.
The final logistic regression model was significant, \( \chi^2(3, N = 31) = 25.48, p < .001 \), indicating that combined performance on the three tasks distinguished adults with SLI from those with TL. The model explained 75\% (Nagelkerke’s \( R^2 \)) of the variance in language status (i.e., affected or unaffected) and correctly classified 87\% of cases, with a sensitivity of 85\% and a specificity of 89\% where cases with a .50 or greater predicted probability were classified as affected. Odds ratios for each of the three standardized variables are listed in Table 5 along with other components of the multitask model. For a one-unit increase in the predictor, an odds ratio of less than 1 indicates how much the odds of being affected decreases (Tabachnick & Fidell, 2007). In the present case, the odds ratios suggested that a one-unit (standard deviation) score increase in any one of the model predictors would result in an approximately equal decline in the likelihood of affectedness—20\% of the prior probability. The standard errors and resulting confidence intervals for the odds ratio estimates, however, were quite large, likely because of the relatively small sample size for this study.

The choice of a .50 or greater predicted probability for identifying affected individuals does not necessarily result in the most accurate classification (Kutner et al., 2005). A review of an ROC curve for the predicted probabilities (see Figure 2) indicated that the optimal predicted probability for the three-variable model was .44. With this new predicted probability rule, the model’s sensitivity improved to 92\% and specificity to 89\%, resulting in an LR+ of 8.3 and an LR– of 0.09. Overall classification accuracy was 90\%, or 28 of 31 participants. The LR+ is above the moderate range (3.0) and short of the extremely positive level (10.0), whereas the LR– is extremely negative based on Dollaghan’s (2007) review of diagnostic evidence. On the whole, the three-predictor cross-task model explained more of the variance in probability of affectedness status and resulted in better likelihood ratios than any of the single-variable models.

### Discussion

The purposes of this study were to explore the effectiveness of three clinical marker tasks for classifying adults with and without SLI and to explore the pattern of adult performance on each of those tasks. Each of the tasks contributed to classification accuracy, suggesting that the processing capacities and linguistic knowledge differences that are assumed to underlie the effectiveness of these tasks for differentiating children with SLI from typical peers are also present in adults. The tasks varied, however, in their classification accuracy. A combination of the three clinical marker tasks resulted in the best overall classification accuracy, with levels of sensitivity and specificity (.92 and .89, respectively) that met Plante and Vance’s (1994) criterion for good discriminant accuracy. The combination of tasks also met Dollaghan’s (2007) suggested level of .10 or below for a negative likelihood ratio. Therefore, the combination of the three tasks showed appropriate accuracy for both ruling in and ruling out a disorder. In contrast, the grammaticality judgments of complex omitted finiteness (COF) sentences.

### Table 5. Final logistic regression model predicting probability that an individual has SLI.

<table>
<thead>
<tr>
<th>Variable</th>
<th>( B )</th>
<th>SE</th>
<th>Odds ratio</th>
<th>90% CI for odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence repetition</td>
<td>–1.66</td>
<td>0.945</td>
<td>0.194</td>
<td>0.04–0.918</td>
</tr>
<tr>
<td>COF A(^a)</td>
<td>–1.55</td>
<td>0.916</td>
<td>0.212</td>
<td>0.05–0.958</td>
</tr>
<tr>
<td>Three-syllable nonword repetition PPC</td>
<td>–1.64</td>
<td>0.898</td>
<td>0.193</td>
<td>0.04–0.846</td>
</tr>
<tr>
<td>Constant</td>
<td>–0.75</td>
<td>0.702</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. All predictors are standardized. CI = confidence interval.

\(^a\)A for grammaticality judgments of complex omitted finiteness (COF) sentences.
for appropriate levels of diagnostic power of LR+ of 10 or higher. Contrary to some prior findings, however, the nonword repetition task results for adults in this study did not have sufficient levels of classification accuracy to meet these clinical guidelines.

**Nonword Repetition**

Dollaghan and Campbell’s (1998) nonword repetition task was not as effective in distinguishing adults as it had been in studies with children. The percentage of three-syllable nonwords repeated correctly was more discriminating than the total phonemes correct score in the context of the combined model, possibly a result of a lack of group differences in one- and two-syllable nonword performance. The greater differences with longer nonwords were consistent with Dollaghan and Campbell’s study with children and several others studies of nonword repetition (Graf Estes et al., 2007). The LR+ for the three-syllable measure was smaller than those based on total phonemes percentage correct found by Dollaghan and Campbell and by Ellis Weismer and colleagues (2000) in studies with children. It is possible that nonword repetition tasks for adults would have greater classification accuracy if they consisted of more items at longer nonword lengths.

Although nonword repetition performance did not meet the standards for a useful diagnostic measure on its own, it did significantly contribute to the classification accuracy of the combined marker model. The findings that nonword repetition performance contributed to classification accuracy of adults with and without SLI was consistent with the findings of Clegg and colleagues (2005), who found deficits on the Nonword Memory Test (Gathercole & Baddeley, 1996) in adults with positive histories of SLI. Clegg and colleagues did not report measures of classification accuracy, only group differences. As in the present study, Clegg and colleagues observed lower performance for the SLI group relative to the typical groups on nonwords longer than three syllables, but not on shorter nonwords.

Several researchers have suggested that the nonword repetition task used in this study taps both segmental phonological analysis ability and phonological working memory (Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Graf Estes et al., 2007). In this study, the SLI group’s performance on forward and backward digit span tasks did not correlate with nonword repetition performance, whereas digits backward did correlate with nonword repetition performance for the TL group. The absence of floor or ceiling effects, and the fact that the correlation trended toward the negative for the affected group, indicates that the results were likely not a statistical artifact. Instead, the results suggest that working memory was less of a factor in nonword repetition performance for adults with SLI and that phonological analysis ability may be a more prominent factor. Similar to Gray’s (2003) findings with young children, forward digit span, a simple memory storage measure, did not significantly correlate with nonword repetition performance for either adult group. Baddeley, Gathercole, and Papagno (1998) argued that their nonword repetition task was a more direct measure of phonological short-term memory than digit span tasks because it was not assisted by the strength of the long-term memory representations of the number words. On the basis of this position, forward digit span measures should not be highly correlated with nonword repetition task performance because they are measuring different constructs.

**Sentence Repetition**

Of the three individual marker tasks, sentence repetition provided the best overall classification accuracy. A reclassification analysis of the SLI group taking into account TOAL–3 sentence imitation performance supported the conclusion that the sentence repetition differences were indicative of a general language ability difference. The two SLI groups resulting from the reclassification did not differ from each other on the CELF–3 Recalling Sentences task but did differ from the TL group.

The ability of sentence repetition to classify adults by their language status was consistent with prior studies of adults with SLI. Sentence repetition contributed to classification accuracy in Tomblin et al.’s (1992) study, and Clegg and colleagues (2005) found that sentence repetition produced a large effect size (d = 3.3) between adults with SLI and those with typical language. Both the Clegg et al. and Tomblin et al. studies used versions of the sentence repetition task from the Multilingual Aphasia Examination, and both studies had participants with a clinical history of language impairment.

Conti-Ramsden and colleagues (2001) found the CELF–3 sentence repetition task to be the best clinical marker of SLI in children age 10 through 11 years. The overall percentage of adults correctly classified by the same task in this study (87%) was nearly identical to the percentage of children correctly classified by the task in Conti-Ramsden et al.’s study. The median raw scores and interquartile ranges for children with SLI were 48 (36–54) and for children with TL 64 (60–68). In the present study, the medians and interquartile ranges on the same measure were 59 (54–62) for adults with SLI and 71 (65–73) for TL adults. The adults with SLI scored better than the children with SLI in Conti-Ramsden et al.’s study, but they did not perform as well as the children with TL. Therefore, although there is evidence of better performance for an older age group, it appears that performance for individuals with SLI may plateau
on this task well below the level achieved by individuals with TL.

Sentence repetition is thought to depend on short-term memory as well as linguistic knowledge, in particular grammatical knowledge (Conti-Ramsden et al., 2001). The magnitude of the correlations of both forward and backward digit span with performance on sentence repetition in this study supports short-term and working memory as factors contributing to individual differences in both the TL group and the SLI group (but note that the correlations for the SLI group did not reach significance). The moderate size of the associations suggests that factors in addition to memory, such as the grammatical knowledge differences between the groups indicated by the grammaticality judgment task performance, may also be influential in adult SLI performance on this task.

**Grammaticality Judgment**

For the grammaticality judgment task, the largest positive likelihood ratio for SLI versus TL group performance was on complex sentences with omitted finiteness markers. The fact that sentences with omitted finiteness were able to discriminate groups is consistent with research conducted by Rice and Wexler (2001) and Conti-Ramsden and colleagues (Conti-Ramsden, 2003; Conti-Ramsden et al., 2001), who found high sensitivity and specificity values for tasks measuring finiteness judgments and use. Because all participants, both SLI and TL, performed well on the complex omitted progressive condition, the overall pattern of results is not consistent with an interpretation that differences are related to attention or metalinguistic awareness. It also does not support a conclusion that the SLI group was unable to be as accurate on sentences simply on the basis of their difficulty processing subject–object relative clauses: The complex omitted finiteness sentences and the complex omitted progressive sentences had identical relative clauses. The group differences are more consistent with differences in grammatical processing or grammatical knowledge.

The pattern of results in this study is largely consistent with the extended optional infinitive account of SLI proposed by Rice, Wexler, and colleagues (Rice et al., 1995, 1996; Rice & Wexler, 1996). Adults with SLI were less likely to detect omissions of finiteness markers and equally likely to detect a nonfiniteness–marking morpheme, progressive–ing when compared with TL peers. Although not statistically significant, the lower level of sensitivity of adults with SLI to sentences with overt agreement errors was not consistent with the extended optional infinitive account and leaves open the question of whether the deficits in grammatical knowledge of adults with SLI are confined to difficulties with grammatical omissions or also include difficulties with overt errors.

One previous study that found that children with SLI were below ceiling levels in judging overt agreement errors was conducted by Wulfeck, Bates, Krupakwoksi, and Saltzman (2004). For the oldest children in the study (i.e., 11–12 years), the group with SLI was less accurate than their typical peers in detecting both errors of overt agreement on be auxiliary verbs as well as errors of omission of be auxiliary verbs. The grammaticality judgments in Wulfeck et al.’s study took place under timed conditions. Miller, Leonard, and Finneran (2008) found that 16-year-olds with SLI were less sensitive than typical peers to overt errors of past tense –ed and third person singular –s. These grammaticality judgments were also obtained under timed conditions. The pattern of errors in these studies suggests that difficulties detecting overt errors of agreement may result from an interaction of processing demands and grammatical knowledge, but the relative contribution of these components to the performance of adults with SLI remains to be explored.

**General Conclusions**

This study found that adults with SLI performed more poorly than TL peers on a range of language-related tasks suggesting that language differences in SLI may be observed into adulthood. With only one exception, the adults in this study who showed evidence of language impairment had no positive history of diagnosis or treatment for language disorders. Several adults in both groups reported a history of speech-language pathology intervention for articulation difficulties. These findings are consistent with those of Zhang and Tomblin (2000) as well as Johnson et al. (1999), who found that children with speech difficulties were more likely to receive intervention than children with language difficulties. Eight of the 13 individuals with SLI who participated in this study reported a history of academic difficulties, primarily reading, suggesting that the absence of intervention cannot be attributed to the presence of mild impairments with no functional impact. More research is needed to build a more complete understanding of the impacts of SLI in adulthood on educational attainment and readiness for employment.

One step toward improving the rate of intervention for young adults with SLI is the development of better assessment methods, including clinical marker tasks. The argument for marker tasks such as nonword repetition and grammaticality judgments focused on finiteness marking is that they aim to identify individuals affected with SLI by focusing on behaviors characteristic of the disorder (Conti-Ramsden, 2003; Rice & Wexler,
Bishop (2006) argued that measures of short-term phonological memory and grammatical morphology may characterize the phenotype of SLI. This study has taken an initial step toward identifying tasks constituting to identify adults with SLI; certainly key refinements remain.

One of these refinements may be to recognize that SLI likely does not result from the presence of a weakness in phonological short-term memory or grammatical morphology alone but is multifactorial (Bishop, 2006). The ideal diagnostic method may be one in which several factors known to be characteristic of SLI are combined into a series of informative measures. The sentence repetition task in this study, as in prior studies with both adults and children, showed the most immediate promise as an individual clinical marker task. The effectiveness of the task may lie in its multifactorial nature (e.g., memory, grammatical competence), although the specific makeup of the factors remains unknown. The effectiveness of the combined tasks provides further evidence that the most promising classification accuracy may come from multifactorial measures. The combined model in this study resulted in odds ratios that were roughly equivalent for each of the standardized predictors, indicating that the probability of being affected with SLI changes an equivalent amount for a unit change in the standardized measure. Furthermore, each of the three predictors contributed to classification accuracy in the presence of the others. Both of these facts indicate that consideration of multiple factors in the SLI group performance made the model more predictive.

The use of clinical marker tasks found to be diagnostic of SLI may prove superior to omnibus language tests both because they are more specific to SLI and because they are a more valid mechanism of classification. Barry et al. (2007), for example, suggested that the large proportion (16%–25%) of control group adults identified by Plante et al. (1996) as affected with SLI was due to the fact that the assessment battery included general literacy measures and so was not sufficiently specific to the identification of SLI. Some clinical marker tasks, such as those based on finite verb morphology and sentence repetition, have been shown to be specific to SLI in children and not a manifestation of developmental disorders in general (Redmond, 2005; Rice, 2003), although there is conflicting evidence in this regard (Eadie, Fey, Douglas, & Parsons, 2002). Further, the use of tasks shown to be characteristic of SLI for assessment of SLI has appeal in terms of the face validity of the measures.

A limitation of this study is the relatively large standard errors of the odds ratios in the combined marker task model. As Dollaghan (2007) pointed out, a measure’s viability for clinical adoption should consider the precision of the likelihood ratios associated with the measure, which relates to the level of certainty that the diagnostic accuracy predicted by the analysis will apply in similar cases in the future. In addition to the refinement of the tasks themselves to improve diagnostic accuracy, the measures require testing with larger samples to reduce standard errors of measure, to determine more precisely their predictive power, and to validate the power of the combined marker model to predict the affectedness of new cases (Kutner et al., 2005).

The evidence from this study and from prior studies with adults builds a compelling case that for many individuals SLI remains unresolved in adulthood (Clegg et al., 2005; Johnson et al., 1999; Plante et al., 1996; Tomblin et al., 1992). The language deficits in adults with SLI result in poorer performance on nonword repetition, sentence repetition, and grammaticality judgment tasks focused on finiteness marking. These performance differences suggest that the deficits underlying SLI in adulthood largely parallel those found in childhood: phonological processing, working memory, and grammatical knowledge of tense and agreement marking. SLI can be identified in adulthood in the absence of a history of diagnosed language impairment, and evidence from this study and others suggests that language impairments remain underdetected and undertreated (Johnson et al., 1999; Zhang & Tomblin, 2000). Further research is required to refine the clinical marker tasks in this study to more precisely estimate the effects of a diagnostic measure combining clinical marker tasks. Research is also needed to better specify the nature of the grammatical deficit associated with SLI in adulthood and to better understand the impact that SLI has on the lives of affected individuals.

Acknowledgments
This work was supported, in part, by funding provided to Gerard H. Poll from the Department of Communication Sciences and Disorders at The Pennsylvania State University. Portions of this article are based on a master’s thesis submitted to the Department of Communication Sciences and Disorders at The Pennsylvania State University by Gerard H. Poll. Preliminary results were presented at the Symposium on Research in Child Language Disorders in Madison, Wisconsin, June 2007. We thank the school staff members who cooperated with this research and the young adults who participated.

References


Received January 21, 2008
Revision received January 9, 2009
Accepted July 1, 2009
DOI: 10.1044/1092-4388(2009/08-0016)

Contact author: Gerard H. Poll, Department of Communication Sciences and Disorders, The Pennsylvania State University, 308G Ford Building, University Park, PA 16802.

E-mail: ghp110@psu.edu.
Appendix. Stimulus items for the grammaticality judgment task.

The cat is pretty and shy.
*The boy smart and funny.
The students called the teacher.
The bird that they fed is blue and yellow.
*The dog that she lost are big and angry.
*The girl that they met short and funny.
The writers listened to rock bands.
*The cow that he saw is walk in the field.
*The bee that she caught are yellow and black.
The players made most of their shots.
*The man that they brought happy and loud.
The drivers complained about the winning car.
Bob and Pat took the driver to the race.
*The kid that they hit are quiet and tall.
*The bee that she caught is crawl up the wall.
*The kid quiet and tall.
The faculty taught English courses in State College.
The boy is smart and funny.
*The cat that he bought are pretty and shy.
The dog that she lost is big and angry.
*The man that they brought are happy and loud.
The discount stores sold calculators.
The cat that he bought is pretty and shy.
*The dog big and angry.
The raccoons scratched the door.
*The kid that they hit is stand in the hall.
*The bird that they fed blue and yellow.
The crows and hawks fought with each other.
The girl that they met is short and funny.
*The cat pretty and shy.
*The cow that he saw spotted and brown.
The kids ate cookies with milk.
The workers stocked the shelves.
*The boy that they chose smart and funny.
The mouse that he heard is hungry and fast.
*The kid that they hit quiet and tall.
*The mouse that he heard are hungry and fast.
The doctors talked about the patient.
The boy that they chose is smart and funny.
The cat that he bought is sleeping next to me.
*The cat that he bought pretty and shy.
The kid is quiet and tall.
*The dog that she lost big and angry.
The players worked hard in spring training.
The kid that they hit is standing in the hall.
*The girl that they met are short and funny.
The dog is big and angry.
*The bird that they fed are blue and yellow.
The bee that she caught is crawling up the wall.
*The cow that he saw are spotted and brown.
The men cleared the snow before ten o’clock.
*The bee yellow and black.
The dog that she lost is barking in the street.
The kid that they hit is quiet and tall.
The cooks made dinner for the group.
*The dog that she lost is bark in the street.
The actors practiced their lines.
*The cat that he bought is sleep next to me.
The grocery stores sold cereal at high prices.
We rode our bikes into town for ice cream.
*The bee that she caught yellow and black.
The teams played good defense and beat the Lions.
The cow that he saw is spotted and brown.
*The boy that they chose are smart and funny.
The bee that she caught is yellow and black.
The carpenters built houses with vinyl siding.
*The mouse that he heard hungry and fast.
The bee is yellow and black.
The man that they brought is happy and loud.
The cow that he saw is walking in the field.
Identification of Clinical Markers of Specific Language Impairment in Adults

Gerard H. Poll, Stacy K. Betz, and Carol A. Miller

*J Speech Lang Hear Res* 2010;53;414-429
DOI: 10.1044/1092-4388(2009/08-0016)

The references for this article include 15 HighWire-hosted articles which you can access for free at: [http://jslhr.asha.org/cgi/content/full/53/2/414#BIBL](http://jslhr.asha.org/cgi/content/full/53/2/414#BIBL)

This article has been cited by 2 HighWire-hosted article(s) which you can access for free at: [http://jslhr.asha.org/cgi/content/full/53/2/414#otherarticles](http://jslhr.asha.org/cgi/content/full/53/2/414#otherarticles)

This information is current as of September 14, 2012

This article, along with updated information and services, is located on the World Wide Web at: [http://jslhr.asha.org/cgi/content/full/53/2/414](http://jslhr.asha.org/cgi/content/full/53/2/414)