Published in final edited form as:

J Speech Lang Hear Res. 2013 December ; 56(6): 1813–1823. doi:10.1044/1092-4388(2013/12-0056).

Identification of Specific Language Impairment in Bilingual Children: I. Assessment in English

Ronald B. Gillam^a, Elizabeth D. Peña^b, Lisa M. Bedore^b, Thomas M. Bohman^b, and Anita Mendez-Perez^b

^aUtah State University, Logan

^bUniversity of Texas at Austin

Abstract

Purpose—This study was designed to derive cut scores for English testing for use in identifying specific language impairment (SLI) in bilingual children who were learning English as a second language.

Method—In a 1-gate design, 167 children received comprehensive language assessments in English and Spanish during their first-grade year. The reference standard was identification by a team of expert bilingual speech-language pathologists. Receiver operating curve (ROC) analyses were used to identify the optimal prediction model for SLI.

Results—The original, English EpiSLI criteria (Tomblin, Records, & Zhang, 1996) yielded a sensitivity of .95 and a specificity of .45 (LR+ = 1.73, LR– = 0.11, and AUC = .79) for our bilinguals. Revised cutoff scores yielded a sensitivity of .86 and a specificity of .68 (LR+ = 2.67, LR– = 0.21, and AUC = .77). An optimal prediction model yielded a sensitivity of .81 and a specificity of .81 (LR+ = 4.37, LR– = 0.23 and AUC = .85).

Conclusion—The results of English testing could be used to make a reasonably accurate diagnostic decision for bilingual children who had attended public school for at least 1 year and were using English at least 30% of the time.

Keywords

assessment; bilingualism; children; language disorders; specific language impairment; primary language impairments

Diagnosing specific language impairment (SLI) in children is difficult under the best of circumstances. The most common practice is to base diagnostic decisions on low scores on standardized language measures. Unfortunately, publishers of standardized tests rarely provide empirically based cutoff scores that yield the highest levels of *sensitivity* (correct identification of children with SLI) and *specificity* (correct identification of typically developing children). In a review of 43 commercially available child language tests, Spaulding, Plante, and Farinella (2006) found that many publishers recommended arbitrary

Correspondence to Ron Gillam: ron.gillam@usu.edu.

Disclosure: Ronald B. Gillam has a financial interest in the Test of Narrative Language (Gillam & Pearson, 2004).

cutoff scores that were likely to identify typical children as impaired. Spaulding et al. (2006) concluded that there is a need for more research that derives empirically based cutoff scores for identifying SLI in specific populations of children.

Perhaps the most influential study of the diagnosis of SLI was a large-scale epidemiological study by Tomblin, Records, and Zhang (1996). Their diagnostic criteria, known as the *EpiSLI model*, was based on an analysis of five composite scores (expressive, receptive, vocabulary, grammar, and narration) that were derived from performance on the Test of Language Development—Primary (Newcomer & Hammill, 1997) and a narrative comprehension and production screening task (Culatta, Page, & Ellis, 1983). For a sample of 1,502 children, the criterion of -1.25 standard deviations below the mean on two or more composites resulted in a sensitivity of .77 and a specificity of .91 (Tomblin et al., 1996), demonstrating the validity and reliability of this model for identifying SLI in monolingual children.

Much less is known about the identification of SLI in bilingual children. Bilingual children with SLI often present difficulties in their first and second languages (L1 and L2, respectively; Peña, Iglesias, & Lidz, 2001; Restrepo & Kruth, 2000). Therefore, assessment in bilingual children's L1 and L2 is the preferred practice (Bedore & Peña, 2008; Gillam, Peña, and Miller, 1999). The general idea is that bilingual children should present poor performance on measures of L1 and L2.

There is some early evidence of the diagnostic accuracy of Spanish and English language measures for identifying SLI in Spanish-English bilinguals. In a recent meta-analysis of this literature, Dollaghan and Horner (2011) calculated the sensitivity, specificity, likelihood ratios, and 95% confidence intervals (CIs) for 17 index measures reported in nine studies. The average LR+ value, which represents the confidence of using scores below the cut-point to diagnose SLI in bilinguals, was 4.12, 95% CIs [2.94, 5.78]. The average LR– value, which represents the confidence in using scores above the cut-point for identifying typical language, was 0.22, 95% CIs [.11, .46]. Dollaghan and Horner concluded that no language measure in Spanish or English stood out as optimal for discriminating between bilingual children with SLI or typical language, with most measures yielding only somewhat suggestive results.

English is a common language for second-language-learning students from a wide range of language backgrounds (Paradis, Schneider, & Duncan, 2012) and for speech-language pathologists (SLPs), who often administer English language standardized assessment instruments to bilingual children in order to inform their diagnostic decisions (e.g., Caesar & Kohler, 2007; Williams & McLeod, 2012).

The current study was designed to derive empirically supported cutoff scores for identifying SLI in bilingual (Spanish-English-speaking) children based on their English language performance. Starting with a group of English-language learners (ELLs) who share Spanish as their first language was a logical first step in the process of examining the diagnostic value of English language testing for the larger population of ELLs in U.S. schools. Studying the diagnostic accuracy of English language testing with Spanish-speaking

children would allow us to use what we know about bilingual development in Spanish and English to independently verify the language learning status of children and then test the diagnostic accuracy of English language testing.

Diagnostic Accuracy of Language Measures With Bilinguals

Examination of bilingual children with SLI suggests that their second language can be used as an indicator of impairment to some extent. In the area of grammatical morphology, Spanish-English and French-English bilinguals have demonstrated English error patterns that were similar to those of monolingual English speakers (Gutiérrez-Clellen & Simón-Cereijido, 2007). In a follow-up study, Gutiérrez-Clellen, Simón-Cereijido, and Wagner (2008) compared bilingual Spanish-English speakers and monolingual English speakers with and without SLI on measures of verb marking and subject use. Monolingual and bilingual children with SLI scored significantly lower than their typical peers, and there were no significant effects associated with bilingual or monolingual status. In addition, similar language patterns have been observed in German monolingual and Turkish-German bilingual children (Rothweiler, Chilla, & Clahsen, 2012) as well as monolingual Dutch and Frisian-Dutch bilingual children (Spoelman & Bol, 2012).

Cross-language comparisons of monolingual and bilingual children do not always yield clear patterns of language similarities. In a study of risk for SLI in Spanish-English preschool-age bilinguals, Peña, Gillam, Bedore, and Bohman (2011) found that English-dominant bilinguals performed similarly to English monolinguals on tests of morphosyntax and semantics that were administered in English, and Spanish dominant bilinguals performed similarly to Spanish monolinguals on tests that were administered in Spanish. However, balanced bilingual children, who were defined as those hearing and speaking both languages between 40% and 60% of the time, had lower scores on Spanish measures than Spanish monolinguals and lower scores on English measures than English monolinguals. In a different study, Gutiérrez-Clellen et al. (2008) found that ELLs presented some language patterns that were similar to monolingual children with SLI and other language patterns that were similar to monolingual children who were typically developing. For example, a group of ELLs made errors on finite verb use that were similar to the errors made by children with SLI. Whereas the monolingual children with SLI made errors on nominative subject use, the ELLs and the typically developing children did not.

The timing of language assessment relative to the second-language-learning process is likely to affect the diagnostic utility of tests administered in either L1 or L2. It has been suggested that school personnel may adopt a “wait and watch” approach to diagnosis of SLI in bilingual children in which evaluators wait for at least a year (Baker, 2001) or longer (Cummins, 2000; Saunders & O’Brien, 2006) to determine whether bilingual children demonstrate a positive trajectory for acquiring the second language. The basic idea is that bilingual children’s language systems are in a state of flux during the early stages of L2 learning. Children who are adept language learners should learn English relatively quickly when they are provided with adequate communication models and support. However, children with SLI who are not adept language learners are unlikely to benefit as much from the kinds of language models and support that are present in regular learning environments.

To the extent that language impairments manifest in L2, it should be possible to predict which children do and do not have language impairments based on the results of testing in the L2, especially when these results are compared to performance in L1.

In keeping with the wait and watch model, we designed this study to determine the accuracy of using L2 testing to diagnose SLI in Spanish-English bilinguals who were using English regularly for at least 1 year. We began to systematically explore the extent to which the assessment of bilingual children's performance in English can inform the diagnosis of SLI in children who are actively engaged in English-language learning.

To address this goal, we calculated the classification accuracy of English testing against a bilingual reference standard (performance on English and Spanish measures). We began with a valid and reliable model of identification, Tomblin et al.'s (1996) EpiSLI model that was established for a large sample of monolingual English speakers. Expecting that the monolingual EpiSLI criteria would not yield high levels of classification accuracy for bilingual children, we wondered if revising the model by resetting the criterion points could yield acceptable classification levels.

The research questions were as follows:

1. What is the diagnostic accuracy of English assessment using the original EpiSLI model (two or more composite scores that are 1.25 *SDs* or more below the mean) for identifying SLI in bilingual children at first grade?
2. Can the EpiSLI model be revised to yield better diagnostic accuracy of SLI in bilingual children?

Method

Recruitment and Screening

Children from 12 elementary schools within three school districts (two in central Texas and one in northern Utah) that served a large proportion of bilingual Latino children were enrolled prospectively during the prekindergarten screening period. The schools were located in small and mid-sized cities and drew children from low and medium socioeconomic status (SES) housing areas. A total of 1,192 children who spoke Spanish (85% of those invited) returned signed consent forms.

All participants were screened prior to kindergarten with the Bilingual English Spanish Oral Screener (BESOS; Peña, Bedore, Gutierrez-Clellen, Iglesias, & Goldstein, 2010), which contained morphosyntax and semantics subtests in both English and Spanish. More information about the children's language skills and performance on the screener is provided by Bohman, Bedore, Peña, Mendez-Perez, and Gillam (2010) and Peña et al. (2011).

The Longitudinal Portion of the Study

The longitudinal study was conducted on 186 children who met the following criteria: Parents identified them as Hispanic (the ethnicity criteria); parents reported that they had at least 20% combined input and output in English and Spanish (the bilingual exposure

criteria); and the children scored below the 30th percentile on at least two of four screening subtests. The screener was derived from the BESOS (Peña et al., 2010), and included subtests of semantics and/or morphosyntax in Spanish plus semantics and/or morphosyntax in English. This approach allowed possible configurations of two, three, or four subtests below the 30th percentile, which increased the likelihood that children with SLI would be included in the longitudinal study. The range of performance on each individual subtest was from 0 to 100% in the selected children. Follow-up testing included a battery of assessments in English and Spanish. The data for this article were drawn from the testing that was administered when the children were in first grade. By the time the children reached first grade, their language skills on all the measures administered in English and Spanish were distributed across the continuum from very low to very high performance.

All the participants were ELLs before kindergarten. Because language proficiency and language ability are difficult to disambiguate in a sample of children with potential SLI, we determined language dominance on the basis of detailed parent and teacher interviews of children's language exposure and use (Gutierrez-Clellen & Kreiter, 2003). Language exposure, use, and history are highly related to children's language performance (Pearson, Fernández, Lewedeg, & Oller, 1997), but the nature of the relationship varies by domain (Bedore et al., 2012; Bohman et al., 2010). Thus, our approach enabled evaluation of language performance across children relative to their linguistic environment.

The majority of these children (53%) had their first exposure to English after 3 years of age, with most having their first exposure to English at age 4. Only three children were reported to have had their first exposure to English when they began school at age 5. Parent and teacher reports of English input and output indicated that 44% of the children were English dominant (listened to or spoke English between 60% and 80% of the time), 26% of the children were balanced bilinguals (listened to or spoke English between 41% and 59% of the time), and 30% of the children were Spanish dominant (listened to or spoke English between 20% and 40% of the time). Most of the children qualified for free (64.6%) or reduced (13.2%) lunch. Of the children's mothers, the majority (61.7%) had not completed high school; 24.5% had completed high school, and 13.8% had attended college.

Nineteen of the 186 children who qualified for the longitudinal portion of the study were lost to follow-up at first grade, leaving a total of 167 children remaining in the study at first grade. The demographics for these 167 students and the 19 students who were lost to follow-up are reported in Table 1. Chi-square tests and *t* tests were used to evaluate the potential significance of the differences between participant characteristics for children with complete data and those who were lost to follow-up. Results showed there were no statistically significant differences between the two groups.

Measures

The EpiSLI identification system (Tomblin et al., 1996) consisted of a two-dimensional matrix of five language composite scores that cross language modality (comprehension and production) with language domain (vocabulary, syntax, and narration). To obtain the data to create the five composites, we administered the Test of Language Development—Primary: 3rd Edition (TOLD-P:3; Newcomer & Hammill, 1997) and the Test of Narrative Language

(TNL; Gillam & Pearson, 2004). The TOLD–P:3 subtests that were administered included Picture Vocabulary, Relational Vocabulary, Oral Vocabulary, Grammatical Understanding, Sentence Imitation, and Grammatical Completion. The TOLD–P:3 was normed on 1,000 children between the ages of 4;0 and 8;11 (years; months). There were 101 Latino children in the TOLD–P:3 norming sample (10%).

The TNL included three narrative formats: stories with no picture cues, stories that corresponded to sequenced pictures, and stories that corresponded with a single picture of a scene. In each condition, children heard a model story and then answered comprehension questions about it. Then, children told their own story. All stories were audio-recorded and transcribed. The TNL was normed on 1,059 children between the ages of 5;0 and 11;11. There were 129 Latino children in the norming sample (12%).

The standard scores from the TOLD–P:3 and TNL subtests were combined to create five composite scores, each with a mean of 10 and a standard deviation of 3. The vocabulary composite was the average of the subtest standard scores from the TOLD–P:3 Picture Vocabulary, Relational Vocabulary, and Oral Vocabulary subtests. The grammar composite was the average of the subtest standard scores from the TOLD–P:3 Grammatical Understanding, Sentence Imitation, and Grammatical Completion subtests. The narrative composite was the average of the standard scores from the TNL Narrative Comprehension and Oral Narration scales. The comprehension composite was the average of the subtest standard scores from the TOLD–P:3 Picture Vocabulary and Grammatical Understanding subtests and the TNL Narrative Comprehension scale. The production composite was the average of the subtest standard scores from the TOLD–P:3 Relational Vocabulary, Oral Vocabulary, Sentence Imitation, and Grammatical Completion subtests and the TNL Oral Narration scale.

Reference Standard

Like Tomblin et al. (1996), we created a reference standard by asking examiners to complete a severity rating scale at the time of testing (see also Records & Tomblin, 1994). Three bilingual SLPs, each with more than 10 years of experience diagnosing and treating SLI in bilingual children, reviewed English and Spanish language samples and the raw scores of testing administered in English and Spanish. The experts independently rated children's performance in the areas of vocabulary, morphosyntax, and narration in English and Spanish on a six-point scale (0 = *severe/profound impairment*, 1 = *moderate impairment*, 2 = *mild impairment*, 3 = *low normal performance*, 4 = *normal performance*, and 5 = *above normal performance*). Raters used the same scale to assign a summary rating for each domain in English and Spanish, as well as an overall summary rating in each language. A child was identified as language impaired when two or three raters assigned a summary rating of 2 or below for each language. To minimize threats to validity from ascertainment bias, the diagnostic status of both affected and unaffected participants was established by the same procedures and reference standards.

Using this system, 21 of the 167 children (12.5%) in the longitudinal sample were identified as presenting SLI in both languages. The overall percentage agreement among the expert bilingual clinicians was 90% (475 agreements out of 498 potential agreements across all

raters). The AC₁ statistic (Gwet, 2008) was used to calculate interrater agreement correcting for chance due to concerns about kappa being highly sensitive to population trait prevalence and differences in rater marginal probabilities (Cicchetti & Feinstein, 1990; Gwet, 2008). The overall AC₁ statistic is defined as

$$AC_1 = \frac{P_a - P_e(\gamma)}{1 - P_e(\gamma)}, \quad (1)$$

with P_a defined as overall agreement probability, $P_e(\gamma)$ defined as $2P_+(1-P_+)$, and P_+ defined as $[A_+ + B_+]/2$, with A_+ and B_+ representing the marginal totals for when the raters identified a student as language impaired. The overall AC₁ statistic was .87, indicating high levels of agreement across the three raters.

Analyses

Receiver operating characteristic (ROC) graphs and logistic regression analyses were used to examine relationships between the EpiSLI diagnostic criteria (performance on comprehension, production, semantics, syntax, and narration composites) and language impairment status. Diagnostic accuracy was assessed by calculating *sensitivity*, defined as the proportion of students that were identified as SLI by the experts and were classified as SLI based on test scores, and *specificity*, defined as the proportion of students that were identified as not having SLI by the experts and who were classified as not impaired based on their test scores. The relationships between sensitivity and specificity can be expressed as likelihood ratios. The positive likelihood ratio (LR+) indicates how likely it is for a test result to occur *below* the criterion value in children who do and do not have SLI. Higher LR+ values correspond to greater confidence that a score below the criterion point can be used to rule in impairment. LR+ values at or above 10.0 indicate that examiners can be very sure that a score below the criterion really comes from a child with SLI. LR+ values around 3.0 indicate that examiners can be moderately sure that scores below the criterion point come from children who really have impairments.

The negative likelihood ratio (LR-) indicates how likely it is for a test result to occur *above* the criterion value in children who do or do not have SLI. Lower LR- values correspond to greater confidence that a score above the criterion point can be used to rule out the presence of impairment. LR- values around 0.1 indicate that examiners can be very sure that a score above the criterion really comes from a child who does not have impairment. LR- values around 0.3 indicate that examiners can be moderately sure that scores above the criterion point come from children who do not have impairments (Dollaghan, 2007; Dollaghan & Horner, 2011).

We also calculated area under the curve (AUC) values, which provide a summary of prediction of true clinical status. The AUC represents the average sensitivity over the range of specificities. An AUC of .5 would indicate a diagnostic marker with no predictive value; an AUC between .8 and .9 would indicate that the diagnostic marker has moderately good

predictive value; and an AUC of 1.0 would indicate that the diagnostic marker has perfect predictive value.

Analyses of ROC graphs were performed to determine the best possible threshold scores. The ROC analysis assessed the full range of sensitivity and specificity at each possible criterion score, which allowed for a full assessment of the potential diagnostic accuracy of the language measures. The optimal threshold was selected by choosing the score that made the binary prediction as close to a perfect predictor as possible, which was the point at which the relationships between the predictors came closest to the perfect curve (Gonen, 2007; Pepe, 2003).

Finally, multivariate logistic regression was used to estimate the joint AUC for the language measures. Logistic regression used a logit transformation of the dichotomous dependent variable (impaired vs. nonimpaired status) to estimate the linear relationships between the language measures and the conditional probability of being impaired versus nonimpaired (Hosmer & Lemeshow, 2000). In this study, π = Probability (Language Status = Impaired|x), where x denotes the vector of explanatory TOLD-P:3/TNL composite scores. We used an iterative method known as *maximum likelihood* to obtain the intercept and coefficient values because this approach did not assume normally distributed outcomes. The additional ROC analysis identified optimal criterion scores based on the predicted probability of being language impaired (e.g., Catts et al., 2001). Individual predicted probabilities were calculated using the following formula: Predicted Probability (0–1 scale) = $1 / \{1 + \text{Exponential} [-(\text{Intercept} + b_1 \times \text{Predictor} + b_2 \times \text{Predictor 2} + b_3 \times \text{Predictor 3} + b_4 \times \text{Predictor 4})]\}$. In this formula, the equation derived from the logistic regression was used to calculate the expected score, which was logit-transformed. This value was then exponentiated and back-transformed into a predicted probability.

Results

Research Question 1

To determine the sensitivity and specificity of the original EpiSLI criteria applied to the bilingual children in our sample, we cross-tabulated language impairment status according to the EpiSLI criteria against independent diagnosis by the bilingual SLPs. Recall that the independent determination of SLI was based on formal and informal language assessments of English and Spanish administered at first grade. The original, English EpiSLI criteria (standard score of 6.25 or less on two or more of the language composites) yielded a sensitivity of .95 and a specificity of .45, resulting in $LR+ = 1.73$, $LR- = 0.11$, and $AUC = .79$. The original EpiSLI criteria for identifying language impairment that was established for monolingual English-speaking children yielded a high true positive rate in our bilingual sample. That is, 20 of the 21 children who were determined to have SLI by expert bilingual clinicians scored below the criterion value of 6.25 on two or more language composites. However, 80 of the 145 children who did not have SLI also scored in the impaired range. The low specificity rate combined with the relatively poor positive likelihood ratio suggests that the original EpiSLI criteria were uninformative for diagnosing SLI in first-grade bilingual children.

Research Question 2

We attempted to improve on the diagnostic accuracy of the EpiSLI model by setting different cutoff scores according to ROC analyses. ROC analyses using logistic regression demonstrated that all five composite dimensions and summary scores were significantly related to the reference standard. Figure 1 shows the ROC curves and AUC values for the subtests and the five composite scores.

Additional ROC analyses were conducted to determine which cutoff scores provided the best balance of sensitivity and specificity for the overall standard score and the five TOLD-P:3/TNL composite scores at first grade (Table 2). Applying the EpiSLI criteria of two composite scores below the new criterion values listed in Table 2 yielded a sensitivity of .86 and a specificity of .68 (LR+ = 2.67, LR- = 0.21, and AUC = .77). The sensitivity value was somewhat better than that for the original EpiSLI study of monolingual English speakers (.86 vs. .77), but the specificity value associated with our revised criterion scores remained far below that of the original EpiSLI study of monolingual children (.68 vs. .91). The 95% CIs for the LR+ [1.97, 3.57] and LR- [0.07, 0.60] indicated that, like the original EpiSLI criteria, our new criteria were minimally useful for diagnosing SLI in bilingual children. The LR+ and LR- 95% CIs for the original and revised EpiSLI criteria overlapped, indicating no significant differences between the two approaches.

We wondered whether any of the seven subtests, acting alone, would yield acceptable levels of specificity and sensitivity. The optimal cut scores, specificity and sensitivity values, and AUC for the seven subtests are listed in the bottom half of Table 2. The Oral Vocabulary subtest on the TOLD-P:3 yielded the highest specificity (.83) but had a sensitivity of only .67. The Grammatical Understanding and Sentence Imitation subtests on the TOLD-P:3 yielded the highest sensitivity (.90) but had specificity values of only .69 and .58, respectively. Most bilingual children with SLI performed poorly on these grammatical measures, but so did many bilingual children without impairments. The Oral Narration scale on the TNL yielded the best relationship between specificity (.78) and sensitivity (.86), with an AUC of .77. None of the subtests, acting alone, yielded minimally acceptable levels of specificity and sensitivity.

Finally, multivariate logistic regression was used to create an optimal prediction model based on all five TOLD-P:3/TNL composite scores. Table 3 shows the estimated regression coefficients that were used in scoring the test results and the outcomes. The equation is presented in the Appendix. For the criterion score (predicted probability) of .18, the optimal prediction model yielded a sensitivity of .81 and a specificity of .81 (LR+ = 4.37, LR- = 0.23, and AUC = .85). The 95% CIs of these LR+ (2.94, 6.52; LR- = .096, 0.57), and the AUC of .85 indicate moderate levels of precision. The LR+ of 4.37 for the optimal prediction model is above the upper limit of the 95% CI of the LR+ (3.57) for the revised EpiSLI model, indicating that the optimal model is superior to the revised EpiSLI model for ruling in SLI. The LR- of 0.23 was within the 95% CI of the LR- for the revised EpiSLI model, indicating that the optimal prediction model and the revised EpiSLI models were of similar use for ruling out SLI. Therefore, a positive result (a test score below the criterion value of .18) is somewhat likely to come from a child with SLI and a negative test score

(above the criterion value) is somewhat likely to come from a bilingual child who does not have an impairment.

Discussion

The identification of SLI in bilinguals is one area in which health disparities across different segments of the population continue to exist. It is difficult to differentiate between bilingual children with true SLI and children who make language errors that are a normal consequence of being in the early stages of second-language learning. Many of the developmental errors children make in their second language are related to knowledge of their first language (Marinis & Chondrogianni, 2011). As their second language gradually becomes more complex, development in the first language may appear to stall or even regress (Anderson, 1999; Kohnert, 2008). This pattern of language development has led to over-identification of SLI in bilingual children and the provision of unnecessary services (Artiles & Bal, 2008). Unfortunately, in an attempt not to overidentify impairments in this situation, clinicians sometimes withhold or delay services when they are actually needed (Bedore & Peña, 2008; Broomfield & Dodd, 2004; Kritikos, 2003). There is a need for reliable methods to appropriately identify SLI in children who are in the process of learning English as a second language.

The primary purpose of this one-gate diagnostic accuracy study was to derive evidence-based cut scores for English tests administered to bilingual, Spanish-English children. We began by determining the extent to which the EpiSLI model that has been validated for identifying SLI in monolingual, English-speaking children (Tomblin et al., 1996, 1997) could be applied to young bilingual children who were learning English. We conducted a series of analyses to evaluate the utility of English language testing for identifying SLI in Spanish-English bilinguals in first grade. The analyses were conducted in a three-step process: (a) calculate sensitivity and specificity using the original EpiSLI criteria; (b) reset the EpiSLI criteria using ROC analyses; and (c) create an optimal prediction model that included children's scores on all the English subtests. Consistent with the "wait and watch" model of diagnosis (Baker, 2001), our analyses were based on English testing that was conducted after all the children had attended public school for at least 1 year.

The Diagnostic Value of the Monolingual EpiSLI Model

The first research goal was to determine the diagnostic accuracy of the cut scores associated with the EpiSLI model that was validated on monolingual English-speaking children when applied to bilingual children. In Tomblin et al.'s (1996) epidemiological study of monolingual English-speaking children, the EpiSLI criteria of two composite scores at or below -1.25 *SD* yielded a sensitivity of .77 and a specificity of .91 ($LR+ = 8.55$ and $LR- = 0.25$) for monolingual English-speaking children.

The original EpiSLI criteria yielded a sensitivity of .95 in our study, indicating that nearly all the bilingual children in our sample who were identified as having SLI by our reference standard scored -1.25 *SD* or more below the mean on two or more language composites. However, the specificity of .45 reflected the fact that 80 of the 145 children who were not impaired were incorrectly identified as having SLI. This unacceptably high level of

overidentification indicates that the original EpiSLI criterion was not low enough to rule out SLI in first grade bilingual children who were actually performing in the typical range in both English and Spanish.

Revising the EpiSLI Cut Scores

In an attempt to improve the diagnostic accuracy of the English testing, we used ROC analyses to reset the criterion scores. Applying the EpiSLI model of two or more composite scores below the new cutoff values resulted in a sensitivity of .86 and a specificity of .68 ($LR+ = 2.67$, $LR- = .21$, $AUC = .77$). Resetting the criterion scores resulted in fewer false negatives, but the improvement in specificity came at the cost of reduced sensitivity. Applying the new cut scores, we failed to identify 14% of the bilinguals who actually had SLI. Neither the likelihood ratios nor the AUC values improved appreciably by changing the criterion values. In addition, no single TOLD-P:3 or TNL subtest yielded minimally acceptable sensitivity or specificity values of .80 or above. Changing to the cut scores associated with the EpiSLI model for identifying SLI in monolingual, English-speaking children (Tomblin et al., 1996; Tomblin et al., 1997) did not result in acceptable levels of diagnostic accuracy for identifying SLI in Spanish-English bilingual children.

An Optimal Prediction Model

We created a multivariate logistic regression equation to predict the probability of SLI given a child's criterion scores for each of the TOLD-P:3 and TNL composites. This was done by employing a multivariate logistic regression technique that estimated the linear relationships between the language measures and the conditional probability of being impaired versus not impaired (Hosmer & Lemeshow, 2000).

Clinicians and researchers can use the logistic regression equation in the Appendix to create a spreadsheet that yields a probability value. After administering the TOLD-P:3 and the TNL, the clinician uses the manuals to calculate the expressive composite, comprehension composite, vocabulary composite, and grammatical composite scores for the TOLD-P:3 and the narrative composite score for the TNL. The composite scores (mean of 10 and *SD* of 3) are entered into the respective columns of the spreadsheet. Then, the equation automatically generates a predicted probability value.

ROC analysis revealed that a probability of .18 or above resulted in minimally acceptable levels of sensitivity (.81) and specificity (.81), which yielded an $LR+$ of 4.37, an $LR-$ of .23, and a statistically significant AUC of .85. The $LR+$ for the optimal prediction model exceeded the upper 95% CI of the revised EpiSLI model, indicating that the optimal equation was better than the revised EpiSLI criteria for identifying SLI in first grade bilingual children. The multivariate approach was more accurate because it allowed each individual measure to share its strengths with the other measures in predicting impairment.

Given the $LR+$ value of 4.37, children who obtain probability scores above the criterion of .18 may have SLI. However, the optimal equation would be insufficient to rule in the disorder because nearly one in four first grade children who obtain a probability score above the criterion of .18 would not be impaired. Similarly, children who obtain a probability score

below the cutoff criteria of .18 are likely not have language impairment. Such a score would not definitively rule out SLI because 23% of these children would actually be impaired.

Users may note that, in some cases, an increase in the comprehension composite scores will correspond to a higher (worse) probability outcome. This occurs because the positive coefficient for the comprehension composite reflects a multivariate estimation that takes all other predictor relationships into account. The bilingual children with SLI in this study were likely to have comprehension composite scores that were somewhat higher than their scores on the other composites.

The Influence of Language Exposure

It is possible that children who were at earlier stages of English language development (those who were exposed to English later or were using less English at home and at school) would be those who would be more likely to score in the language-impaired range when they weren't really impaired (the "false positive" group). Similarly, children at the later stages of English language development might be more likely to score in the normal range even though they were really impaired (the "false negative" group).

Examining first exposure to English, we found that the children in the false positive group started learning English later than children in the other three groups. The false positive children were first exposed to English at 2;11 on average, which is 2 months later than the mean for the false negative children, 4 months later than the true positive children, and 7 months later than the mean for the true negative children. In terms of English usage, the false positive children were using less English at home and at school ($M = 39\%$) than the true positive children ($M = 44\%$), the true negative children ($M = 48\%$), or the false negative children ($M = 52\%$). Thus, the children who were at the earliest and latest stages of learning and using English were those who were likely to be misdiagnosed by English-only testing. However, there was 65%–80% overlap in the distributions of first exposure to English and English usage across the children in the true positive, false positive, true negative, and false negative subgroups.

We also examined the relationship between amount of exposure to English and the likelihood of being misclassified. We rank ordered the children in the false positive and false negative groups according to English language exposure. Children with English exposure rates at or below 30% input and output at school and at home were misclassified (false positives and false negatives) 29% of the time, compared to 17% misclassification of children who had exposure rates above 30%. These findings are consistent those of Gutiérrez-Clellen et al. (2008). In that study, the ELL children with less than 30% exposure to English at home were the most likely to be misclassified.

Interpreting the Diagnostic Accuracy of English Assessment of Bilingual Children

Diagnosis of SLI in children is a complex process. Even tests with relatively strong psychometric properties may not discriminate between monolingual children with typical and impaired language very well. Plante and Vance (1994) recommended that clinicians select tests with sensitivity and specificity levels of .80 and above. In their review of commercially available tests of child language, Spaulding et al. (2006) found sensitivity and

specificity values were not available or could not be calculated for most of the 43 tests that they reviewed. Of the 18 tests for which diagnostic accuracy could be calculated, the average sensitivity was .76 and the average specificity was .86 (Spaulding et al., 2006, Table 4). Ten of those tests had sensitivity and specificity values that both exceeded .80 for the identification of SLI in monolingual English speakers (see also Friberg, 2010).

Our optimal equation yielded .81 sensitivity and .81 specificity for identifying SLI in bilingual speakers using only the results from English testing. In the Introduction we raised the idea that the kinds of difficulties children have learning English as a second language fall in line with the difficulties some children have learning English as a first language. If testing assesses those aspects of English that pose difficulties for monolingual and bilingual children with language impairments, it should not be too surprising that English testing could yield similar degrees of diagnostic accuracy for monolingual and bilingual speakers.

Recall that Dollaghan and Horner (2011) conducted a meta-analysis of nine studies of methods for diagnosing language disorders in bilingual Spanish-English children between 3 and 15 years. The pooled LR+ across 17 language measures (13 of which were Spanish) was 4.12, 95% CI [2.94, 5.78]. This compares quite closely to the positive likelihood ratio for our optimal equation (LR+ = 4.37, 95% CI = 2.94 – 6.52). The pooled LR– value from the meta-analysis was .22, 95% CI [.11, .46], which is quite similar to our optimal equation LR– value of .23, 95% CI [.096, .57]. All nine studies that made it into the Dollaghan and Horner meta-analysis employed two-gate designs in which children who were known to have SLI were compared to children who were known to be typically developing. We used a type of a one-gate design in which we did not know which children presented SLI until the end of the study. The diagnostic classifications were not known to any of the individuals who administered or scored the tests. This is relevant because the diagnostic accuracy of a test is susceptible to inflation in two-gate designs. Given these two factors, our results suggest that English testing has the potential to yield diagnostic accuracy metrics for bilingual children that are reasonably similar to metrics for testing in Spanish.

Many of the bilingual children in the United States speak languages for which bilingual testing protocols do not exist. For these children, English testing may take on added importance. As recommended by Spaulding et al. (2006), empirically derived cut scores are needed for a variety of populations that are tested by SLPs. Our classification accuracy for testing in English was similar to the .83 sensitivity and .80 specificity on English grammatical testing reported by Gutiérrez-Clellen and Simón-Cereijido (2007). If test cutoffs are appropriately set for children who are in the process of acquiring English (via norms or via resetting cut points), examiners can be successful in ruling in and ruling out SLI.

Of course, bilingual children must have some degree of English use in order for this approach to work. In the Gutiérrez-Clellen and Simón-Cereijido (2007) study, the children used English 50% of the time or more. In the current study, all the bilingual children used English at least 20% of time. Analysis of the students who were misclassified suggested that the results of English testing could be used to make a fairly accurate diagnostic decision for

children who had attended public school for at least 1 year and were using English at least 30% of the time.

Potential Limitations

The participants in this study scored below the 30th percentile on one or more subtests in English and Spanish that were administered 2 years before the diagnostic accuracy data were collected. We reasoned that bilingual children whose performance was in the low-typical range or whose performance by domain and language was inconsistent (e.g., high Spanish morphosyntax and English semantics but low Spanish semantics) relative to monolingual children in each language would be challenging to distinguish from bilinguals with true SLI. It is possible that our diagnostic accuracy metrics may have been deflated because the longitudinal portion of our study reported herein did not include the full range of performance on the four subtests of the screener given in English and Spanish before kindergarten. However, over the 2-year period in which data were collected after screening, the participants dispersed along the affected-to-unaffected continuum. By the time they were tested during their first-grade year, the bilingual children in our sample fell at all points along the oral language continua in both English and Spanish on all of the administered measures. Therefore, we do not believe that the diagnostic accuracy rates reported in this article were deflated.

The results of this study could have been partially affected by *incorporation bias*, which occurs when the diagnostic test under consideration is used as the reference standard. In this study, the results from the TOLD-P:3 and the TNL were reviewed by a committee of three expert bilingual SLPs in order to determine which participants had SLI. Results of Spanish language testing, language sample data in English and Spanish, and case history information were available to the expert clinicians, who independently applied their own criteria to identify the combination of measures that differentiated between the participants who were or were not language impaired. In our opinion, basing diagnostic decisions on the larger set of Spanish and English data should have reduced the likelihood of incorporation bias. Nonetheless, it is possible that some degree of incorporation bias could have inflated our estimates of the accuracy of the TOLD-P:3 and TNL measures.

Our research design also failed to completely control for subjective bias. Two of the raters had also tested some of the participants. However, at least two of the three raters did not test each participant and had no opportunity for contact. As with incorporation bias, subjective bias would tend to inflate the sensitivity and specificity values. This threat to internal validity was partially minimized because participants were identified as language impaired only if two of three raters agreed.

Clinically, even the regression equation that yielded the best LR+ and LR– values should be used with caution. According to our data, we can assume that the pretest probability of SLI in children who scored below the 30th percentile on two of four subtests (one in each language) is 12.5%. Given the regression equation LR+ value of 4.37, the posttest probability is only 45% that a child who scores at or above the criterion probability value will actually have SLI. But, given the LR– value of 0.23, it is only 5% likely that a child who obtains a probability value below the criterion actually has the disorder. Therefore, as

has been argued by Kohnert (2010) and Kohnert, Windsor, and Yim (2006), English testing may be more useful to rule out SLI in bilingual children than to rule it in. Ruling in SLI with high degrees of certainty may require testing in both L1 and L2. We are currently analyzing data to test this hypothesis.

Conclusion

The identification of SLI in bilinguals who are in the process of learning English is quite difficult. Assessment in bilingual children's first and second language is the preferred practice (Bedore & Peña, 2008; Gillam et al., 1999). Unfortunately, the usefulness of bilingual assessment is limited by the combination of numerous languages being spoken in school settings, few bilingual SLPs, lack of developmental data on many languages, few tests that are normed on bilinguals, and a lack of empirically derived cut scores for determining good versus poor performance in either L1 or L2. English is the common language for all bilingual, second-language learners in the United States, and English language testing routinely informs diagnostic decisions about ELLs at risk for SLI (e.g., Caesar & Kohler, 2007; Williams & McLeod, 2012). This study provides evidence-based criteria for using English language testing to identify language impairments in bilingual children who have been enrolled in school for at least 1 academic year and are listening to and speaking English at least 30% of the day. Knowledge of the evidence supporting the best possible cut scores combined with information from observations and the nature of parent or teacher concerns (Paradis et al., 2012) should serve to improve the accuracy of diagnostic decisions.

Acknowledgments

This research was supported by National Institute on Deafness and Other Communication Disorders Grant R01DC007439.

We thank all of the interviewers and testers for their assistance with collecting the data and the school districts for allowing us access to the participants.

References

- Anderson R. 1999; Impact of first language loss on grammar in a bilingual child. *Communication Disorders Quarterly*. 21:4–16.
- Artiles AJ, Bal A. 2008; The next generation of disproportionality research: Toward a comparative model in the study of the equity in ability differences. *Journal of Special Education*. 42:4–14.
- Baker, C. *Fundamentals of bilingual education and bilingualism*. Buffalo, NY: Multilingual Matters; 2001.
- Bedore LM, Peña ED. 2008; Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*. 11:1–29.
- Bedore LM, Peña ED, Summers C, Boerger K, Resendiz M, Greene K, ... Gillam RB. 2012; The measure matters: Language dominance profiles across measures in Spanish/English bilingual children. *Bilingualism: Language and Cognition*. 15:616–629.
- Bohman TM, Bedore LM, Peña ED, Mendez-Perez A, Gillam RB. 2010; What you hear and what you say: Language performance in early sequential Spanish English bilinguals. *International Journal of Bilingual Education and Bilingualism*. 13:325–344. DOI: 10.1080/13670050903342019 [PubMed: 21731899]

- Broomfield J, Dodd B. 2004; The nature of referred subtypes of primary speech disability. *Child Language Teaching and Therapy*. 20:135.
- Caesar LG, Kohler PD. 2007; The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools*. 38(3):190–200.
- Catts HW, Fey ME, Zhang X, Tomblin JB. 2001; Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools*. 32:38–50.
- Cicchetti DV, Feinstein AR. 1990; High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*. 43:551–558. [PubMed: 2189948]
- Culatta B, Page JL, Ellis J. 1983; Story retelling as a communicative performance screening tool. *Language, Speech, and Hearing Services in Schools*. 14:66–74.
- Cummins, J. *Language, power and pedagogy: Bilingual children in the crossfire*. Clevendon, United Kingdom: Multilingual Matters; 2000.
- Dollaghan, CA. *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Brookes; 2007.
- Dollaghan CA, Horner EA. 2011; Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*. 54:1077–1088.
- Friberg JC. 2010; Considerations for test selection: How do validity and reliability impact diagnostic decisions? *Child Language Teaching and Therapy*. 26:77–92. DOI: 10.1177/0265659009349972
- Gillam, RB, Pearson, N. *Test of Narrative Language*. Austin, TX: Pro-Ed; 2004.
- Gillam RB, Peña ED, Miller L. 1999; Dynamic assessment of narrative and expository discourse. *Topics in Language Disorders*. 20:33–47.
- Gonen, M. *Analyzing receiver operating characteristic curves with SAS*. Cary, NC: SAS Institute, Inc; 2007.
- Gutiérrez-Clellen VF, Kreiter J. 2003; Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*. 24:267–288.
- Gutiérrez-Clellen VF, Simón-Cerejido G. 2007; The discriminant accuracy of a grammatical measure with Latino English-speaking children. *Journal of Speech, Language, and Hearing Research*. 50:968–981.
- Gutiérrez-Clellen VF, Simón-Cerejido G, Wagner C. 2008; Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Applied Psycholinguistics*. 29:3–19. [PubMed: 22685359]
- Gwet K. 2008; Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology*. 61:29–48. [PubMed: 18482474]
- Hosmer, DW, Lemeshow, S. *Applied logistic regression*. 2. New York, NY: Wiley; 2000.
- Kohnert, K, editor. *Language disorders in bilingual children and adults*. San Diego, CA: Plural; 2008.
- Kohnert K. 2010; Bilingual children with specific language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders*. 43:456–473. DOI: 10.1016/j.jcomdis.2010.02.002 [PubMed: 20371080]
- Kohnert K, Windsor J, Yim L. 2006; Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research & Practice*. 21:19–29.
- Kritikos E. 2003; Speech-language pathologists' beliefs about language assessment of bilingual/bicultural individuals. *American Journal of Speech-Language Pathology*. 12:73–91. [PubMed: 12680815]
- Marinis T, Chondrogianni V. 2011; Comprehension of reflexives and pronouns in sequential bilingual children: Do they pattern similarly to L1 children, L2 adults, or children with specific language impairment? *Journal of Neurolinguistics*. 24:202–212.
- Newcomer, PL, Hammill, DD. *Test of Language Development—Primary*. 3. Austin, TX: Pro-Ed; 1997.
- Paradis J, Schneider P, Duncan TS. 2012; Discriminating children with language impairment among English Language Learners from diverse first language backgrounds. *Journal of Speech, Language, and Hearing Research*. doi: 10.1044/1092-4388(2012/12-0050)

- Pearson BZ, Fernández SC, Lewedeg V, Oller DK. 1997; The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*. 18:41–58.
- Peña ED, Bedore LM, Gutiérrez-Clellen VF, Iglesias A, Goldstein B. 2010 Bilingual English Spanish Oral Language Screener. Experimental Version.
- Peña ED, Gillam RB, Bedore LM, Bohman TM. 2011; Risk for poor performance on a language screening measure of bilingual preschoolers and kindergarteners. *American Journal of Speech-Language Pathology*. 20:302–314. [PubMed: 21821821]
- Peña ED, Iglesias A, Lidz CS. 2001; Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology*. 10:138–154.
- Pepe, MS. The statistical evaluation of medical tests for classification and prediction. New York, NY: Oxford University Press; 2003.
- Plante E, Vance R. 1994; Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in the Schools*. 25:15–24.
- Records NL, Tomblin JB. 1994; Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *Journal of Speech and Hearing Research*. 37:144–156. [PubMed: 8170120]
- Restrepo MA, Kruth K. 2000; Grammatical characteristics of a Spanish-English bilingual child with specific language impairment. *Communication Disorders Quarterly*. 21:66–76.
- Rothweiler M, Chilla S, Clahsen H. 2012; Subject–verb agreement in specific language impairment: A study of monolingual and bilingual German-speaking children. *Bilingualism: Language and Cognition*. 15(1):39–57. DOI: 10.1017/s136672891100037x
- Saunders, W, O'Brien, G. Oral language. In: Genesee, F, Lindholm-Leary, K, Saunders, W, Christian, D, editors. *Educating English language learners: A synthesis of empirical evidence*. Cambridge, United Kingdom: Cambridge University Press; 2006. 24–97.
- Spaulding TJ, Plante E, Farinella KA. 2006; Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in the Schools*. 37:61–72.
- Spoelman M, Bol GW. 2012; The use of subject–verb agreement and verb argument structure in monolingual and bilingual children with specific language impairment. *Clinical Linguistics & Phonetics*. 26(4):357–379. DOI: 10.3109/02699206.2011.637658 [PubMed: 22404865]
- Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. 1997; Prevalence of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*. 40:1245–1260.
- Tomblin JB, Records NL, Zhang X. 1996; A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research*. 39:1284–1294. [PubMed: 8959613]
- Williams CJ, McLeod S. 2012; Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology*. 14(3):292–305. DOI: 10.3109/17549507.2011.636071 [PubMed: 22472031]

Appendix. Creating a Spreadsheet for Generating Predicted Probabilities of SLI

SLPs can use the table below to create a spreadsheet that will generate the predicted probabilities of SLI. Five TOLD–P:3 and TNL composite scores should be entered into a spreadsheet that is capable of handling an exponential function (EXP in the formula below). The formula for each time point will need to be altered to refer to the column names (and appropriate cell) that are noted by the letters enclosed in parentheses in column headings. The formula should be copied and pasted into Column H with the revisions just described. Entering the composite values with the appropriate equation (e.g., First Grade Predicted Probability Equation) will provide the predicted probability value, as follows:

$$\text{Predicted probability} = 1 / \{1 + \text{EXP}[-(1.8191 + -0.2183 \times \text{Expressive Composite} + 0.0534 \times \text{Comprehension Composite} + 0.1778 \times \text{Vocabulary Composite} + -0.2633 \times \text{Grammatical Composite} + -0.4618 \times \text{Narrative Composite})]\}$$

Table A1 presents the predicted probabilities of SLI for two students' TOLD-P:3/TNL composite scores. The criterion score was .18. Students who obtain predicted probabilities that are greater than the criterion score would be identified as language impaired. Students who obtain predicted probabilities less than the criterion score would be identified as not language impaired. In the example below, Student 1 would not be considered to have SLI but Student 2 would.

Table A1

Predicted probabilities of specific language impairment (SLI) for two students' Test of Language—Primary: 3rd Edition (TOLD-P:3) and Test of Narrative Language (TNL) composite scores.

Student (A)	Time (B)	TOLD-P:3/TNL Composite Scores					Probability of SLI (H)
		Exp (C)	Comp (D)	Vocab (E)	Gram (F)	Narrative (G)	
1	First	9.5	15.7	11.5	11.3	14.0	.0011
2	First	3.8	7.3	6.5	4.7	5.0	.2667

Note. Exp = expressive composite; Comp = comprehension composite; Vocab = vocabulary composite; Gram = grammatical composite; Narrative = narrative composite.

Figure 1.

Receiver operating characteristic (ROC) curves and area under the curve (AUC) values for different Test of Language—Primary: 3rd Edition (TOLD-P:3) and Test of Narrative Language (TNL) composites at first grade.

Table 1

Comparison of participants with complete data and those who were lost to follow-up.

Participant Characteristic	Complete data			Lost to follow-up		
	<i>n</i>	%	<i>M</i>	<i>SD</i>	<i>n</i>	%
Gender						
Female	83	50			7	37
Male	84	50			12	63
School district						
Texas 1	52	31			6	32
Texas 2	72	43			5	26
Utah	43	26			8	42
Mother's education						
< 7th grade	52	31			7	37
7th–11th	51	31			8	42
HS grad	41	25			0	0
Partial college	14	8			3	16
College grad or grad school	9	5			1	5
Age in months			63.2	4.3		64.6
Year of first exposure to English			2.1	1.9		2.8
						1.8

Table 2

Specificity and sensitivity estimates for total score, the five composite scores, and the seven subtest scores.

Diagnostic marker	Optimal threshold	Specificity	Sensitivity	AUC
Total score	69	.71	.67	.75
Expressive composite	4.5	.78	.71	.81
Comprehension composite	6.67	.74	.81	.78
Vocabulary composite	5.5	.77	.57	.66
Grammatical composite	5.0	.73	.81	.80
Narrative composite	6	.73	.91	.83
TOLD Grammatical Understanding	8	.69	.90	.75
TOLD Oral Vocabulary	5	.83	.67	.75
TOLD Picture Vocabulary	7	.67	.67	.56
TOLD Relational Vocabulary	4	.78	.67	.67
TOLD Sentence Imitation	4	.58	.90	.73
TNL Narrative Comprehension	6	.76	.86	.82
TNL Oral Narration	6	.78	.86	.77

Note. AUC = area under the curve value; TOLD = Test of Language Development—Primary: 3rd Edition; TNL = Test of Narrative Development.

Table 3

Multivariate logistic regression prediction of SLI using the five TOLD–P:3/TNL composite measures (Panel A) and the resulting diagnostic accuracy (Panel B).

A		
Parameter	First grade	
Intercept	1.82	
Grammatical composite	−0.26	
Expressive composite	−0.22	
Narrative composite	−0.46	
Comprehension composite	0.05	
Vocabulary composite	0.18	
AUC	0.85	

B		
Multivariate LR prediction	Reference standard	
	SLI	NLI
SLI	17	27
NLI	4	119
Total agreement	21 (0.81)	146 (0.81)

Note. LR = likelihood ratio; SLI = specific language impairment; NLI = not language impaired.