



Published in final edited form as:

Int J Speech Lang Pathol. 2020 April ; 22(2): 152–162. doi:10.1080/17549507.2019.1632931.

Screening school-age children for developmental language disorder in primary care

Kerry Danahy Ebert, PhD,

Department of Communication Disorders and Sciences, Rush University, Chicago, IL

Cesar Ochoa-Lubinoff, MD,

Department of Pediatrics, Rush University Medical Center, Chicago, IL

Melissa P. Holmes, MD

Department of Pediatrics, Rush University Medical Center, Chicago, IL.

Abstract

Purpose: To evaluate the feasibility, preliminary diagnostic accuracy, and reliability of a screening tool for Developmental Language Disorder (DLD) in early school-age children seen in a pediatric primary care setting.

Method: Sixty-six children aged 6- to 8-years attending well-child visits at a large urban pediatric clinic participated. Parents completed a 5-item questionnaire and children completed a 10-item sentence repetition task. A subset of participants ($n = 25$) completed diagnostic testing for DLD. Exploratory cut-offs were developed for the parent questionnaire, the child sentence repetition task, and the combined score.

Result: The screening tool could be reliably implemented in two minutes by personnel without specialty training. The best diagnostic accuracy measures were obtained by combining the parent questionnaire and child sentence repetition task. The tool showed strong internal consistency, but the parent and child scores showed only moderate agreement.

Conclusion: The screening tool is promising for utilisation in primary care clinical settings, but should first be validated in larger and more diverse samples. Both the parent and child components of the screening contributed to the preliminary findings of high sensitivity and specificity found in this study. Screening for DLD in school age children can increase awareness of an under-recognised disorder.

Keywords

identification; diagnostic accuracy; developmental disorders; pediatricians

Developmental Language Disorder (DLD) (Bishop et al., 2017) affects approximately 7 to 10 percent of children (Johnson et al., 1999; Tomblin, 1997). DLD has also been called Specific Language Impairment and Primary Language Impairment, and is captured as mixed

CORRESPONDING AUTHOR: Kerry Danahy Ebert, Kerry_ebert@rush.edu, 312-563-6034 (ph), 312-942-7211 (fax); 1016D AAC, 600 S. Paulina St., Chicago, IL 60612.

AUTHOR DISCLOSURE STATEMENT: The authors declare no conflict of interest.

receptive-expressive language disorder (F80.2) or expressive language disorder (F80.1) in ICD-10. Here we follow the recent recommendation of the international CATALISE consortium (Bishop, Snowling, Thompson, Greenhalgh, & the CATALISE-2 Consortium, 2017) and adopt the term DLD. Children with DLD demonstrate clinical deficits in language skills, such as vocabulary and grammar, in comparison to unaffected peers (Leonard, 2014; Paul & Norbury, 2012). Deficits commonly persist across time, even into early adulthood (Johnson, 1999). DLD is typically diagnosed via direct assessment of children's language skills (Paul & Norbury, 2012; Tomblin et al., 1997). Parental checklists or symptom reports have been developed to complement direct assessment (Norbury, Nash, Baird, & Bishop, 2004; Paradis, Emmerzael, & Sorenson Duncan, 2010), but are not typically used in isolation.

During the school years, DLD has pervasive negative social and academic effects. Children with the disorder develop friendships of lower quality (Durkin & Conti – Ramsden, 2007), experience more peer rejection (Fujiki, Brinton, Hart, & Fitzgerald, 1999) and demonstrate higher levels of social withdrawal (Hart, Fujiki, Brinton, & Hart, 2004). They are also at higher risk of emotional and behavioural disorders (Brownie et al., 2004; Zadeh, Im- Bolter, & Cohen, 2007). Academically, children with DLD demonstrate persistently lower reading skills than unaffected peers (Catts, Bridges, Little, & Tomblin, 2008). Longitudinal studies indicate that children affected by DLD achieve lower levels of education and select occupations with lower socioeconomic status than unaffected peers (Johnson, Beitchman, & Brownlie, 2010). It has also been documented that later in childhood and adolescence, children with DLD more frequently experience emotional, behavioural, and attention-deficit/hyperactivity problems and have more severe forms of these disorders than children with typical language abilities (Yew & O'Kearney, 2013).

However, identification of DLD is poor. In a group of 216 kindergarten children diagnosed with DLD in an epidemiological study, just 29% of parents reported ever having been informed that their child had a speech or language problem (Tomblin et al., 1997). More recently, a group of 286 Australian children were screened twice during the kindergarten year by their teachers (Jessup, Ward, Cahill, & Keating, 2008). This procedure identified just 15% of children affected by DLD. In another study conducted in Australian schools (Antoniazzi, Snow, & Dickson-Swift, 2010), 15 teachers provided ratings of children's oral language skills on a standardised instrument. When these ratings were compared to direct testing of the children's oral language, correspondence was poor. Diagnoses derived from the teacher ratings had high rates of both false positives and false negatives (Antoniazzi et al., 2010).

The causes of these poor identification rates are likely multifactorial. Possible contributors include inconsistent terminology and limited professional and public awareness of the nature of language and of language disorders (Bishop et al., 2017; Kamhi, 2004); the association of language disorders with multiple causes and multiple professions (Tomblin et al., 1997), which may obscure the nature of the difficulty and the appropriate referral path when problems are suspected; and the relatively covert nature of impairments in language comprehension, especially compared to behavioural or speech production problems and especially in cases of mild to moderate DLD.

Screening for Developmental Disorders in Pediatric Primary Care

Identification of DLD and other developmental disorders (such as autism spectrum disorder and attention-deficit hyperactivity disorder) is one responsibility of pediatric primary care clinics. The American Academy of Pediatrics (AAP) policy statement in 2002 established that early identification of developmental disorders is critical to the well-being of children and their families. It is an integral function of the primary care medical home and an appropriate responsibility of all pediatric health care professionals (American Academy of Pediatrics, 2002). A revised AAP policy statement (American Academy of Pediatrics, 2006) provided specific guidance on developmental surveillance and screening by primary care providers for children from birth to 36 months of age, including developmental surveillance at all well child visits and structured developmental screenings at 9, 18, and 30 (or 24) months of age. Finally, the AAP also developed the Bright Futures Guidelines for Health Supervision of Infants, Children, and Adolescents (American Academy of Pediatrics, 2006) which are designed to promote overall health across child development. Bright Futures Guidelines recommend screening for developmental concerns, including speech and language, at regular intervals as part of well-child visits. Irrespective of the procedures used, when a child screens positive, the primary care provider should make a referral for further evaluation and treatment.

There is a notable paucity of data on identification rates of developmental-behavioural disorders in primary care (Sheldrick, Merchant, & Perrin, 2011). A systematic review of the limited data available indicated that physicians working without a validated screening tool have difficulty identifying developmental or behavioural problems (Sheldrick et al., 2011). In particular, this scenario resulted in low sensitivity rates (below 54% in almost all studies reviewed). In contrast, the AAP recommends that developmental screening instruments demonstrate sensitivity and specificity rates higher than 70% (Committee on Children with Disabilities, 2001). It is thus important to create and validate developmental screening instruments. For children under five years of age, developmental screening instruments with appropriate levels of sensitivity and specificity are available (Sices, Stancin, Kirchner, & Bauchner, 2009), although the same cannot be said for children 5 years and older.

Despite screening recommendations and existing tools, surveys have repeatedly demonstrated that the majority of physicians do not perform routine screening using standardised tools (Sices, Feudtner, McLaughlin, Drotar, & Williams, 2003; Sand et al., 2005; Limbos, Joyce, & Roberts, 2009). Potential contributors to physician reluctance to screen include inadequate time or remuneration for screening processes, conflicting reports on the accuracy of available screening tools, and the limited research that has been conducted in primary care settings (Sices et al., 2009; Sices et al., 2003; Sand et al., 2005). For speech-language disorders specifically, recent systematic reviews from the US Preventive Services Task Force (USPSTF) (Nelson, Nygren, Walker, & Panoscha, 2006; Berkman et al., 2015) have highlighted both the conflicting data on screening accuracy and the limitations of the research base. Although some screening instruments can accurately identify children with language disorders (Berkman et al., 2015), the USPSTF's overall assessment remains that there is insufficient evidence to recommend for or against routine use of formal language screening instruments in primary care in children up to 5 years of

age (Nelson et al., 2006). The lack of research considering costs and benefits of screening plays a key role in this finding (Wallace et al., 2015) and further research into speech and language screenings is recommended.

The main focus of developmental screening in primary care has been in children under three years of age. Currently there are no AAP recommendations for developmental screening after children turn three years of age and there are no major studies on the use of developmental or behavioural screening tools for children older than five years of age. Although the school system is expected to play a major role in identification of developmental disorders in school-age children, available research suggests that -- at least in the case of DLD -- identification via the school system alone may be inadequate (Jessup et al., 2008). Screenings conducted in pediatric primary care may be able to complement school-based screening in order to improve the overall identification of DLD.

However, primary care screening tools for school-age children must first be developed and investigated. Moreover, the primary barriers to screening implementation in this setting (e.g. time) must be considered in the development of screening tools. To our knowledge, existing screening tools for DLD in school-age children (e.g. Clinical Evaluation of Language Fundamentals – 5th Edition Screener, Wiig, Secord, & Semel, 2013) require 10 minutes or more for completion. With 80% of preventative care visits lasting less than 20 minutes total (Halfon et al., 2011), screening tools of this length are not feasible. In this study, preliminary data for a new screening tool that could be implemented in primary care was collected.

The new screening tool was designed to merge prior research on DLD with feasible primary care screening practices. In primary care, developmental screening is commonly accomplished via parent report alone. However, DLD is most commonly identified via direct child assessment. We decided to include both parent report and child assessment in the screening in order to evaluate the contributions of both components. Although parent report is not commonly included in diagnostic testing for DLD, there is evidence that parent report tools can contribute to accurate identification of the disorder (Norbury et al., 2004; Paradis et al., 2010; Restrepo, 1998) and that they may be combined with child assessment to increase accuracy (Bishop & McDonald, 2009; note that statistical concerns arise in combining child assessment and parent report when they use different measurement scales, a point to which we return in the discussion). For the child assessment component, we selected a sentence repetition task. Poor sentence repetition has been noted to be one of the most promising behavioral markers for DLD (Archibald & Joannis, 2009; Pawlowska, 2014), perhaps because it appears to index underlying language abilities (Klem et al., 2015). Pawlowska (2014) reviewed studies that have investigated the diagnostic accuracy of sentence repetition tasks and found that evidence suggests the task shows moderate accuracy in ruling in or ruling out DLD. However, variability across studies is substantial and further research is needed (Pawlowska, 2014).

The purpose of the current study was to assess the feasibility and accuracy of this new dual-component screening tool for DLD in school-age children seen in the primary care setting. The specific objectives for the study included: (1) to assess the time and personnel required for the screening; (2) to determine preliminary diagnostic accuracy for the tool, including

both the individual parent and child components and the combined score; and (3) to examine the internal consistency of the screening tool, particularly the agreement of parent report and direct child assessment in a school-age population.

METHOD

This study followed a prospective design. The study was approved by the Institutional Research Board at the institution where it was conducted.

Participants

A convenience sample of children was recruited at a large urban pediatric primary care clinic associated with an academic medical center. Charts for scheduled well-child appointments were reviewed and determined to be eligible if they met the following criteria: (1) child aged 6;0 (years;months) through 8;11; (2) child has no documented developmental disabilities (e.g. autism spectrum disorder, hearing or vision impairment, intellectual disability; a complete list was provided to research assistants); (3) parent's preferred language listed as either English or Spanish. The age range was selected to target the early school years and to match the age range for the planned follow-up DLD testing. The final criterion was included because the screening tool was originally designed to be administered to either English- or Spanish-speaking parents; however, all parents who consented were English-speaking (see Result below). Eligible parents were approached and invited to participate and informed consent was obtained from 66 parents. Written assent was obtained from children at least 7 years of age, consistent with institutional policy.

Test Methods

Screening procedures.—For the first 47 children tested, the screening procedures were administered by one of three graduate students with no previous background in speech-language pathology. The remaining 19 children were screened by a speech-language pathology graduate student. For all children, screening was administered in the private clinic exam room while the children were waiting for a provider. Procedures were audio recorded for feasibility and reliability analyses.

The DLD screening test contained two parts: a parent questionnaire and a child sentence repetition task. Parents were given a written questionnaire with 5 questions regarding language development and concerns. Questions were derived from validated parent report tools (Paradis et al., 2010; Restrepo, 1998). More specifically, the 5 questions asked the parent to rate the child's ability to express his or her ideas, the child's ability to comprehend instructions and questions, the child's performance in school, the frequency with which the child produces complex sentences, and the ease with which the child can explain something that happened when the parent was not there. For each question, parents selected a response on a scale from 0 to 3. All points on the response scale were anchored with a description. For example, one question asked parents, "Is it easy for your child to explain or describe things that happened when you were not there?" Response options included: "0 = no, usually hard; 1 = sometimes not easy; 2 = easy enough; 3 = very easy." Another question asked, "How often does your child produce long and complicated sentences (for example,

sentences containing words like because, although, when)?” with response options “0 = less than once a day (or never); 1 = once or twice a day (or occasionally); 2 = several times a day (or frequently); 3 = almost every time he or she speaks (or always).” The total possible score for the parent questionnaire was 15. The complete parent questionnaire is available from the authors.

Children completed a 10-item sentence repetition task constructed by Redmond (2005). Instructions and 1 practice item with feedback were first presented to the child. These were identical to the instructions and practice item in Redmond (2005). Next, 10 test sentences containing 10 words each were presented. The original set of 16 items published in Redmond (2005) was reduced to increase feasibility for the clinical setting; specifically, items 1-4 and 6-11 (p. 127, Redmond, 2005) were presented. As in Redmond (2005), the sentences were an equal mix of active and passive constructions. All instructions and stimuli were presented using live voice. Examiners were instructed not to repeat test items; if the child requested a repetition, the examiner marked the item incorrect and moved to the next sentence.

Children were instructed to repeat each item verbatim. Items were immediately scored as correct or incorrect by the examiner. Any deviation from the original sentence resulted in a score of incorrect. In particular, changes to morphological markers (e.g. omission of *-ed* or *-s* endings) were included as errors. Articulatory errors that did not alter morphology (e.g. distortions of /r/) were not counted as errors. Resulting scores were summed and then weighted by 1.5, such that the total possible score for the child sentence repetition task was 15. Without an apriori reason to weight one screening component more heavily than the other in the composite screening score, this step was taken so that the maximum score of the parent and child components was equal.

DLD diagnostic testing.—Children who completed the screening were invited to participate in a follow-up study on a separate date. The follow-up study (Ebert, Rak, Slawny, & Fogg, 2019) included diagnostic testing for DLD. The individuals who administered the diagnostic tests were unaware of the children’s screening scores. Diagnostic testing included a standardised language test, the Clinical Evaluation of Language Fundamentals – 4th Edition (CELF-4; Semel, Wiig, & Secord, 2003), and a complete parent report tool, the Alberta Language Development Questionnaire (ALDeQ; Paradis et al., 2010). Children also passed a hearing screening and scored within no more than 1.25 standard deviations below the mean on a nonverbal intelligence test. The four subtests of the CELF-4 that make up the Core Language Composite score were administered. These subtests assess a child’s ability to follow directions, produce appropriate grammatical forms, repeat sentences, and construct sentences. Using a cut-off of –1.5 standard deviations below the mean for diagnosing DLD results in 100% sensitivity and 89% specificity for this tool (Semel et al., 2003).

The ALDeQ is a 4-section parent questionnaire designed to assess early language development, current communication skills, and family history of communication disorders (Paradis et al., 2010). In a sample of dual-language learners from minority language households (a population considered difficult to diagnose for DLD) the tool demonstrated 66% sensitivity and 96% specificity (Paradis et al., 2010). The protocol for the diagnostic

testing study (Ebert et al., 2019) used the ALDeQ to corroborate test scores with functional information on the presence or absence of everyday difficulty with language; the ALDeQ was specifically chosen over other parent report tools because it was designed for use with diverse language learners, who were included in the diagnostic testing study. Children who scored in the DLD range on the CELF-4 were expected to also have evidence of parent concern regarding language skills, as evidenced by scores below the suggested cut-off on the ALDeQ. Children who scored above the DLD range on the CELF-4 were expected to have no evidence of parent concern regarding language skills, as evidenced by scores above the suggested cut-off on the ALDeQ. In cases of mismatch between the CELF-4 and the ALDeQ, children's DLD status was considered indeterminate (i.e. they were not counted in analyses). For children who scored within the 90% confidence interval (CI) of the CELF-4 cut-off, the ALDeQ scores were used to differentiate between children with and without DLD.

Analyses.—Audio recordings of the sentence repetition task were used to calculate time for the screening as well as scoring reliability. An independent research assistant trained in speech-language pathology listened to recorded data for all participants and documented the length of the sentence repetition task. The research assistant also rescored each sentence from the recording and calculated point-by-point agreement with the original score.

The accuracy of the screening tool was calculated separately for the parent questionnaire score, the child sentence repetition score, and the total screening score created by summing these two components. Because this was a new screening tool, all cut-offs were exploratory rather than pre-specified. Cut-offs were not set at the time the screening test was scored, and therefore individuals scoring the screener were not aware of the test cut-offs. Cut-offs were subsequently developed using a Receiver Operating Characteristic (ROC) analysis implemented in SPSS version 22. Cut-offs were selected to maximise the combination of sensitivity and specificity.

After the cut-offs were set, diagnostic accuracy measures were calculated. Diagnostic accuracy measures included sensitivity, specificity, and likelihood ratios. Sensitivity (i.e. the screening test's ability to correctly identify children with DLD) and specificity (i.e. the screening test's ability to rule out children without DLD) can be interpreted using the minimum value of .70 recommended by the AAP (Committee on Children with Disabilities, 2001). Alternatively, Plante and Vance (1994) propose a more stringent framework, with sensitivity and specificity values of 0.80-0.89 considered fair, values of 0.90 and higher considered good, and values below 0.80 considered not useful diagnostically. For the interpretation of likelihood ratios, Dollaghan (2007) provides guidelines: for positive likelihood ratios, 1 is a neutral or uninformative test, 3 is a moderately positive test that may be suggestive but insufficient to confirm a diagnosis, and 10 or greater is a very positive test. For negative likelihood ratios, 1 is again a neutral or uninformative test, 0.30 is a moderately negative test that may be suggestive but insufficient to rule out a diagnosis, and 0.10 or smaller is a very negative test (Dollaghan, 2007).

Internal consistency of the screening tool was evaluated using Cronbach's alpha and inter-item correlations. Cronbach's alpha values between .70 and .90 are considered acceptable

(Tavakol & Dennick, 2011). The consistency of results between parents and children was of particular interest and the Pearson correlation between the total scores for each of these two sections was calculated. Pearson correlations were interpreted according to Cohen (1988): $r = .10$ is a small effect, $r = .30$ is a medium effect, and $r = .50$ is a large effect.

RESULT

Participants

The flow of participants through the study is depicted in Figure 1. The parents of 66 children consented to participate. All consents were completed in English, according to expressed parent preference. Of the 66 consents, three children refused to participate in the sentence repetition task. The children that refused the task included one 6-year-old boy and two 7-year-old girls. Data from these children and their parents are excluded from all analyses.

The remaining 63 children included 34 boys and 29 girls. They ranged in age from 6;0 through 8;11 with a mean age of 7;2. Table I shows descriptive statistics for performance on the DLD screening tool.

Because the follow-up diagnostic testing required scheduling and attending a separate appointment, approximately 38% of the original sample (25 of 66 children) completed the diagnostic testing for DLD. The group that completed diagnostic testing included 11 boys and 14 girls; they ranged in age from 6;0 to 8;10, with a mean age of 7;5. Differences between the group that completed follow-up and the group that did not were explored to verify that the two groups did not differ systematically. The group that completed follow-up differed from the group that did not complete follow-up in age ($t(61) = -2.29, p = .026$), but not in gender ($\chi^2(1) = 1.67, p = .198$), scores on the parent questionnaire portion of the screening task ($t(61) = 1.02, p = .311$), or scores on the child sentence repetition portion of the screening task ($t(61) = 0.25, p = .804$).

Of the group that completed follow-up testing, five children met criteria for DLD: two children scored below the 90% CI of the CELF-4 cutoff and also had parent concern on the ALDeQ, and three children scored within the 90% CI and had parent concern. Twenty children in the follow-up sample demonstrated typical language development: three children scored within the 90% CI but had no evidence of parent concern and 17 children scored above the upper bound of the 90% CI cutoff on the CELF-4. There were no cases considered indeterminate in the follow-up sample. Table II displays the characteristics and scores of the groups with and without DLD in the follow-up sample.

Screening Feasibility

The first goal of the study was to assess the time and personnel required to complete the screening task. The parent questionnaire required minimal time or personnel, as it was brief and written. The child sentence repetition task was administered by personnel without specialty training (i.e. graduate students without speech-language training) for the first 47 children. For these three graduate students, reliability analyses indicated that the original administrator and scorer of the task agreed with the expert scorer on 90.9% of sentences. When data from children screened by the speech-language pathology graduate student were

added into reliability analyses, scoring agreement occurred on 90.8% of all sentences. Based on the audio recordings, the sentence repetition task took a mean time of 2 minutes, 8 seconds to administer when all administrators were included, with a range of 1 minute, 8 seconds to 3 minutes, 42 seconds. The task took a mean time of 1 minute, 56 seconds when only the three administrators without speech-language pathology background were included (range 1 minute, 8 seconds to 2 minutes, 47 seconds) and a mean time of 2 minute, 37 seconds (range 1 minute, 35 seconds to 3 minutes, 42 seconds). These results were interpreted to indicate that the screening task could be administered quickly and accurately by individuals without specialty training. Finally, qualitative feedback from clinic personnel indicated that the screening process was minimally disruptive to the clinic workflow.

Screening Accuracy

The second goal of the study was to assess the preliminary diagnostic accuracy of the screening tool. First, cut-offs were developed based on analysis of scores in the subset of children who completed follow-up diagnostic testing. For the total screening tool, setting the cut-off at 14 (i.e. scores of 14 or greater indicated a pass and scores less than 14 indicated failure) optimised the sensitivity/specificity tradeoff. For the parent questionnaire, the cut-off was set at 8 and for the sentence repetition the cut-off was set at 6 (based on the weighted score). Table III shows the cross-tabulation of screening test results by DLD diagnostic testing results for both the individual screening components and the total screening score.

Next, the sensitivity, specificity, positive and negative predictive values, and positive and negative likelihood ratios, along with 95% CIs for each parameter, were calculated for the DLD screening test. These values are displayed in Table IV. Based on the limited number of children in the sample, both the combined tool and the sentence repetition task showed perfect sensitivity. The combined tool also showed good specificity (Plante & Vance, 1994), but the sentence repetition task had specificity at the AAP's minimum of 0.70 (and below the minimum suggested by Plante & Vance, 1994). The parent questionnaire showed perfect specificity, but sensitivity fell below the 0.70 minimum. For the combined tool, both the positive and negative likelihood ratios met the standards for a very informative test (i.e. 10 or greater for the positive ratio, 0.10 or less for the negative ratio; Dollaghan, 2007). However, the CIs for the likelihood ratios show that the true values may fall within a wide range. In the cases of "perfect" likelihood ratios (i.e., negative likelihood ratios of 0.00 for the full screening tool and for the sentence repetition task; positive likelihood ratio of ∞ for the parent questionnaire), the precision of the point estimates cannot be calculated.

Screening Reliability

The final goal of the study was to assess the internal consistency of the screening tool, including the agreement between parent report and child performance. Cronbach's α was .813 for the parent questionnaire alone and .779 for the child sentence repetition task alone. In both of these analyses, deletion of any single item from the screening tool resulted in poorer internal consistency (i.e. lower Cronbach's α score). Internal consistency for the combined screening tool was similar to the components, $\alpha = .800$. All Cronbach's α values (i.e., parent questionnaire, sentence repetition, and combined tool) fell within the acceptable range (Tavakol & Dennick, 2011).

The total score for the parent component correlated significantly with the child sentence repetition score, $r(63) = .293, p = .020$. Figure 2 shows the relationship between child sentence repetition scores and parent questionnaire scores for participants who were diagnosed with DLD, those who were negative for DLD on follow-up testing, and those who did not complete the follow-up diagnostic testing. The figure illustrates the positive linear relationship between the two components. The high sensitivity of the sentence repetition component (using the cut-off of 6) and the high specificity of the parent questionnaire (using the cut-off of 8) are also illustrated.

DISCUSSION

This study presented preliminary data on the implementation of an innovative two-part screening tool for DLD within the primary care setting. Analyses focused on three main questions. The first question concerned the feasibility of the screening procedures within the clinic. The most commonly endorsed barriers to developmental screening among physicians are insufficient time and reimbursement for the activity (e.g. Halfon et al., 2001) and thus it was critical to create a feasible procedure. Results in this area were encouraging. Parents could complete the questionnaire quickly, during wait times. The sentence repetition task required approximately 2 minutes to implement and score. Reliability analyses indicated that graduate students with no prior training in speech-language pathology could administer the task quickly and reliably score the sentences while administering the task.

Based on these results, we anticipate that the screening could be implemented by clinic support personnel without specialty training. These brief procedures could be administered by medical assistants or nurses when obtaining growth measurements and vital signs for well child visits without interfering with clinic patient flow. Sentence repetition stimuli could be recorded to facilitate administration and to eliminate variability in the presentation of the stimuli. Scoring can be done during administration, and it would be straightforward to implement an algorithm to sum scores and provide the screening result to the provider. However, we note that the personnel in the present study were students, not clinic support personnel, and the ability of these personnel to complete the screening should be empirically confirmed.

One potential weakness of the feasibility of the screening procedure was the proportion of children (4.5%) who refused to complete the sentence repetition task. It is possible that fewer children would refuse if the task were administered by familiar clinic personnel in the context of other clinical procedures, rather than unfamiliar research assistants who entered the patient room specifically for this purpose. However, this possibility would have to be explored empirically.

The second question was whether the screening measure, which was created for this study, could accurately identify children with DLD. Diagnostic testing results were available for only a small sample, and therefore results should be interpreted cautiously. However, the tool appears promising. The combined tool achieved perfect sensitivity and good specificity (per Plante & Vance, 1994). Although the CI for the positive likelihood ratio is large, the lower bound (i.e. the worst possible value) falls at Dollaghan's (2007) cut-off for a

moderately informative test. As noted in the Result, the precision of the negative likelihood ratio cannot be calculated in this case.

Neither the parent questionnaire nor the sentence repetition task in isolation achieved the same level of diagnostic accuracy as the combined tool. The parent questionnaire was highly specific but its sensitivity was inadequate. The sentence repetition task was highly sensitive but its specificity was inadequate (per Plante & Vance's 1994 guidelines). In other words, both components of the screening tool were needed to achieve optimal sensitivity and specificity. Of course, it is important to reiterate the limitation of the small sample size in this study, and also to note that the tasks in the diagnostic testing for DLD were similar to the tasks in the screener. In addition, the parent input was obtained via a rating scale (i.e., an ordinal measurement) whereas the child input was obtained via an accuracy count (i.e., ratio measurement). These represent different types of scores and it would be psychometrically preferable to transform the parent questionnaire score via Rasch analysis (Rasch, 1980). The focus of this initial study was clinical feasibility and such an analysis could not be undertaken. Ultimately, the screening tool should be validated within a larger sample before it is implemented clinically, and this step could also involve statistical procedures to explore transforming the parent questionnaire scores.

The final study question related to the internal consistency of the scale and the agreement between the parent and child components. Internal consistency, as measured by Cronbach's α , fell within the acceptable range (Tavakol & Dennick, 2011) for the child sentence repetition task alone, the parent questionnaire alone, and the two components combined. In addition, the parent questionnaire and child sentence repetition task correlated significantly, nearly reaching the medium effect size according to conventions in psychological research. However, this correlation coefficient ($r = .293$) indicates that just 8.6% of the variance overlaps between the two components (i.e. $r^2 = .086$), suggesting that each of the components contributes related but distinct information to the screening tool.

Challenges and Future Directions

There are a number of barriers to screening older children for developmental disorders such as DLD in primary care settings. Successfully implementing a primary care DLD screening tool would likely require prior training for physicians to increase their awareness of the prevalence and impact of this condition in school-age children. Although we are not aware of data on physician awareness of DLD and its impact on other areas of functioning, both clinical experience and data on the rate of identification of DLD (Tomblin et al., 1997) suggest physician awareness is likely low. Having a screening tool that can be used during well child visits could help to maintain physician awareness for this important condition. In addition to awareness of the condition, physicians report unfamiliarity with screening tools and billing codes, as well as a shortage of staff to assist with screenings (Halfon et al., 2001). Shortages of diagnostic and treatment services for patients who fail screenings are also a concern (Halfon et al., 2001), and additional research on the efficacy of these services is needed (Sices et al., 2003).

Yet screening for developmental disorders is an important issue which warrants attention. An estimated 12% to 16% of the U.S. pediatric population has a developmental disorder

(Boyle, Decoufle, & Yeargin-Allsopp, 1994). There has been increasing pressure to identify these children at an earlier age (American Academy of Pediatrics, 2002; 2006). However, at least in the case of DLD, a large proportion of affected children are not identified before school age (Tomblin et al., 1997; Jessup et al., 2008). In addition, disorders such as DLD may not fully manifest before the school years (Johnson et al., 1999; Rudolph & Leonard, 2016).

Primary care physicians have a role in screening for conditions that affect the developmental trajectories of patients and in advocating for appropriate treatment services through schools or other therapy resources. We can identify several needed steps towards this goal in the case of DLD in the school years. First, the tool tested here requires validation within a larger sample. It will also be important to validate the tool across diverse populations. We note the potential of this tool to screen children from minority language households; the parent questionnaire could be translated into the home language, with the sentence repetition task implemented in English, in order to obtain information about development in both languages. This potential is crucial in the linguistically diverse U.S. but was not tested in the current study. It will also be important to consider the expansion of the tool beyond the age range tested in this study. For example, including 5-year-olds could facilitate earlier identification of DLD. Ultimately, it will be important to also increase awareness of DLD and other developmental disorders among primary care physicians and to implement training on validated screening tools.

CONCLUSION

In summary, our paper is an initial effort to investigate the use of an efficient and effective screening tool that could be used to identify a highly prevalent condition that can interfere with the self-esteem, academic achievement and social performance of the children that suffer from it. Additional investigation of school-age screenings for developmental disorders in primary care is needed.

ACKNOWLEDGMENTS:

The authors thank Klaudia Bednarczyk, Tirsit Berhanu, Allison Byrne, Rachel Slager, and Omar Taibah for their assistance with data collection and management.

FUNDING SOURCES: Supported in part by NIH R03DC013760 awarded to K. Ebert

The funding agency had no involvement in the study design, collection, analysis, or interpretation of data, or in the writing of the report or decision to submit for publication.

REFERENCES

- American Academy of Pediatrics. (1990). *Bright Futures: Guidelines for Health Supervision of Infants, Children, and Adolescents* (4th Ed). Elk Grove Village, IL: American Academy of Pediatrics.
- American Academy of Pediatrics, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. (2002). The medical home. *Pediatrics*, 110, 184–186. [PubMed: 12093969]
- American Academy of Pediatrics, Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, and Medical Home Initiatives for Children With Special Needs. (2006). Identifying infants and young children with developmental

disorders in the medical home: an algorithm for developmental surveillance and screening [published correction appears in *Pediatrics*, 118(4), 1808–1809.

- Antoniazzi D, Snow P, & Dickson-Swift V (2010). Teacher identification of children at risk for language impairment in the first year of school. *International journal of speech-language pathology*, 12(3), 244–252. [PubMed: 20433343]
- Archibald LD, & Joanisse MF (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, 52(4), 899–914.
- Berkman ND, Wallace I, Watson L, Coyne-Beasley T, Cullen K, Wood C, & Lohr KN (2015). Screening for speech and language delays and disorders in children age 5 years or younger: A systematic review for the U.S. Preventive Services Task Force Evidence Syntheses, No. 120 Rockville, MD: Agency for Healthcare Research and Quality (US).
- Bishop DV, & McDonald D (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders*, 44(5), 600–615. [PubMed: 19387886]
- Bishop DM, Snowling MJ, Thompson PA, & Greenhalgh T (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology And Psychiatry*, 58(10), 1068–1080. [PubMed: 28369935]
- Boyle CA, Decoufle P, & Yeargin-Allsopp M (1994). Prevalence and health impact of developmental disabilities in US children. *Pediatrics*, 93, 399–403. [PubMed: 7509480]
- Brownlie EB, Beitchman JH, Escobar M, Young A, Atkinson L, Johnson C, & ... Douglas L. (2004). Early language impairment and young adult delinquent and aggressive behavior. *Journal of Abnormal Child Psychology*, 32(4), 453–467. [PubMed: 15305549]
- Catts HW, Bridges MS, Little TD, & Tomblin JB (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research*, 51(6), 1569–1579.
- Cohen J (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Committee on Children with Disabilities. (2001). Developmental surveillance and screening of infants and young children. *Pediatrics*, 108(1), 192–195. [PubMed: 11433077]
- Dollaghan CA (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore, MD: Paul H. Brookes.
- Durkin K, & Conti-Ramsden G (2007). Language, social behavior, and the quality of friendships in adolescents with and without a history of specific language impairment. *Child Development*, 78(5), 1441–1457. [PubMed: 17883441]
- Ebert KD, Rak D, Slawny CM, & Fogg L (2019). Attention in bilingual children with Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, 62, 979–992.
- Fujiki M, Brinton B, Hart CH, & Fitzgerald AH (1999). Peer acceptance and friendship in children with specific language impairment. *Topics in Language Disorders*, 19(2), 34–48.
- Halfon N, Hochstein M, Sareen H, O'Connor KG, Inkelas M, & Olson LM (2001, 4). Barriers to the provision of developmental assessments during pediatric health supervision. In *Pediatric Academic Societies Meeting*.
- Halfon N, Stevens GD, Larson K, & Olson LM (2011). Duration of a well-child visit: association with content, family-centeredness, and satisfaction. *Pediatrics*, peds-2011.
- Hart KI, Fujiki M, Brinton B, & Hart CH (2004). The relationship between social behavior and severity of language impairment. *Journal of Speech, Language, and Hearing Research*, 47(3), 647–662.
- Jessup B, Ward E, Cahill L, & Keating D (2008). Teacher identification of speech and language impairment in kindergarten students using the Kindergarten Development Check. *International Journal of Speech-Language Pathology*, 10(6), 449–459. [PubMed: 20840024]
- Johnson CJ, Beitchman JH, Young A, Escobar M, Atkinson L, Wilson B, & ... Wang M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research*, 42(3), 744–760.

- Johnson CJ, Beitchman JH, & Brownlie EB (2010). Twenty-year follow-up of children with and without speech-language impairments: Family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology*, 19(1), 51–65. [PubMed: 19644128]
- Kamhi AG (2004). A meme's eye view of speech-language pathology. *Language, Speech, and Hearing Services in Schools*, 35(2), 105–111.
- Klem M, Melby-Lervåg M, Hagtvet B, Lyster SAH, Gustafsson JE, & Hulme C (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, 18(1), 146–154. [PubMed: 24986395]
- Leonard LB (2014). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.
- Limbos MM, Joyce DP, Roberts GJ (2009). Nipissing District Developmental Screen: Patterns of use by physicians in Ontario. *Canadian Family Physician*, 56, 66–72.
- Nelson HD, Nygren P, Walker M, & Panoscha R (2006). Screening for speech and language delay in preschool children: Systematic evidence review for the US Preventive Services Task Force. *Pediatrics*, 117(2), 298–319.
- Norbury CF, Nash M, Baird G, & Bishop DM (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: A validation of the Children's Communication Checklist-2. *International Journal of Language & Communication Disorders*, 39(3), 345–364. [PubMed: 15204445]
- Paradis J, Emmerzael K, & Duncan TS (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, 43(6), 474–497. Questionnaire available at: <https://cloudfront.ualberta.ca/-/media/arts/departments-institutes-and-centres/linguistics/chesl/documents/aldeq.pdf> [PubMed: 20304411]
- Paul R & Norbury C (2012). *Language Disorders From Infancy Through Adolescence*. St. Louis, Mo.: Elsevier.
- Pawłowska M (2014). Evaluation of three proposed markers for language impairment in English: A meta-analysis of diagnostic accuracy studies. *Journal of Speech, Language, and Hearing Research*, 57(6), 2261–2273.
- Plante E, & Vance R (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, 25(1), 15–24.
- Rasch G *Probabilistic Models for Some Intelligence and Attainment Tests*. The University of Chicago Press, IL, USA (1980).
- Redmond SM (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics & Phonetics*, 19(2), 109–127. [PubMed: 15704501]
- Restrepo MA (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, 41(6), 1398–1411.
- Rudolph JM, & Leonard LB (2016). Early language milestones and specific language impairment. *Journal of Early Intervention*, 38(1), 41–58.
- Sand N, Silverstein M, Glascoe FP, Gupta VB, Tonniges TP, & O'Connor KG (2005). Pediatricians' reported practices regarding developmental screening: Do guidelines work? Do they help? *Pediatrics*, 116, 174–179. [PubMed: 15995049]
- Semel E, Wiig EH, & Secord WA (2003). *Clinical evaluation of language fundamentals* (4th ed.). San Antonio, TX: The Psychological Corporation.
- Sheldrick RC, Merchant S, & Perrin EC (2011). Identification of developmental-behavioral problems in primary care: A systematic review. *Pediatrics*, 128(2), 356–363. [PubMed: 21727101]
- Sices L, Feundtner C, McLaughlin J, Drotar D, & Williams M (2003). How do primary care physicians identify young children with developmental delays? A national survey. *Journal of Developmental and Behavioral Pediatrics*, 24(6), 409–417. [PubMed: 14671474]
- Sices L, Stancin T, Kirchner L, & Bauchner H (2009). PEDS and ASQ developmental screening tests may not identify the same children. *Pediatrics*, 124, 640–647.
- Tavakol M & Dennick R (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. [PubMed: 28029643]
- 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 & Hearing Research*, 40(6), 1245–1260.

- Wallace IF, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, & Lohr KN (2015). Screening for speech and language delay in children 5 years old and younger: A systematic review. *Pediatrics*, 136(2), e449–e462.
- Wiig EH, Secord WA, & Semel E (2013). *Clinical evaluation of language fundamentals* (5th ed.): Screening test. San Antonio, TX: The Psychological Corporation.
- Yew SK, & O’Kearney R (2013). Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies. *Journal of Child Psychology and Psychiatry*, 54(5), 516–524. [PubMed: 23082773]
- Zadeh ZY, Im-Bolter N, & Cohen NJ (2007). Social cognition and externalizing psychopathology: An investigation of the mediating role of language. *Journal of Abnormal Child Psychology*, 35(2), 141–152. [PubMed: 17191148]

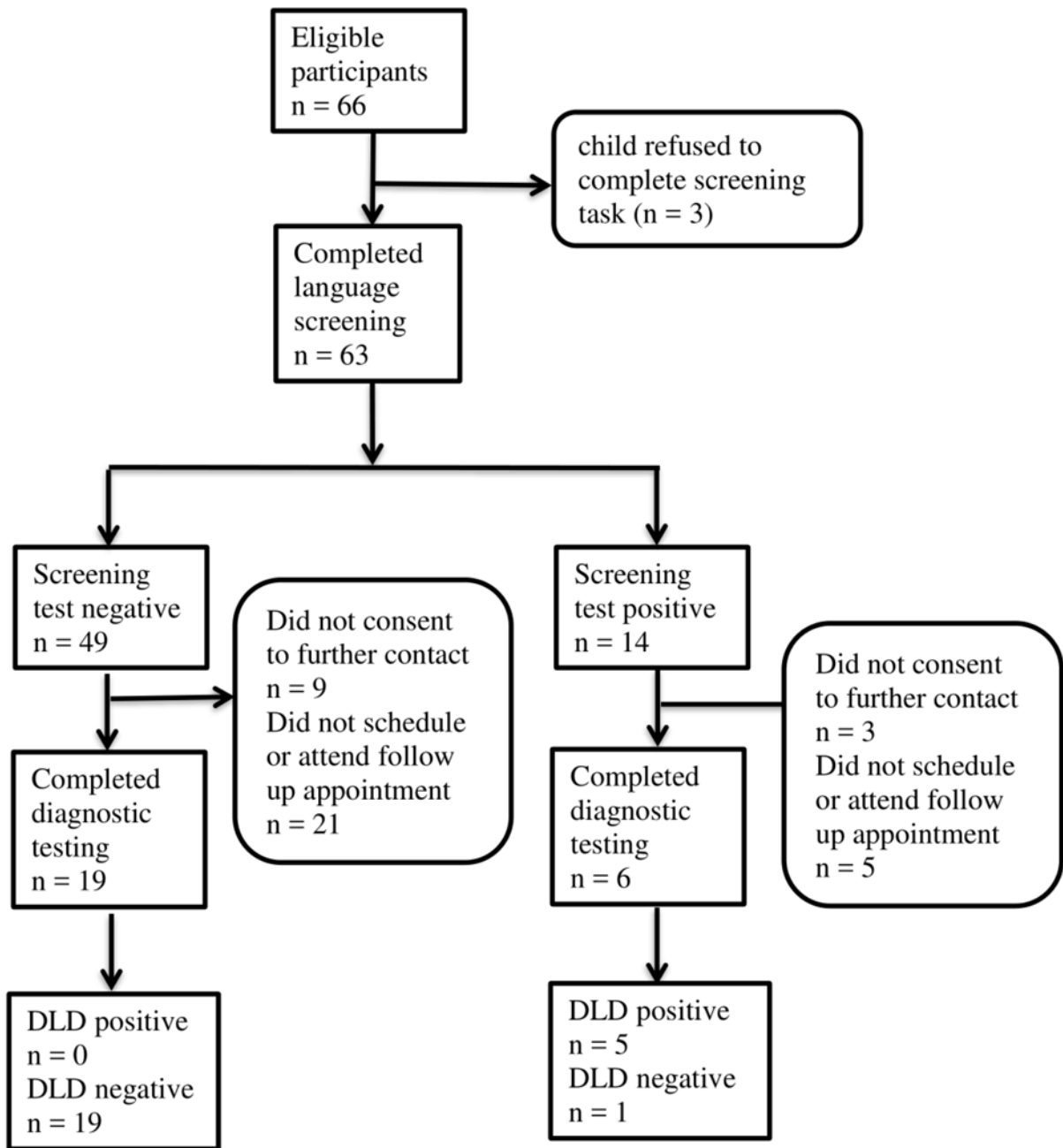


Figure 1.
Participant flowchart.

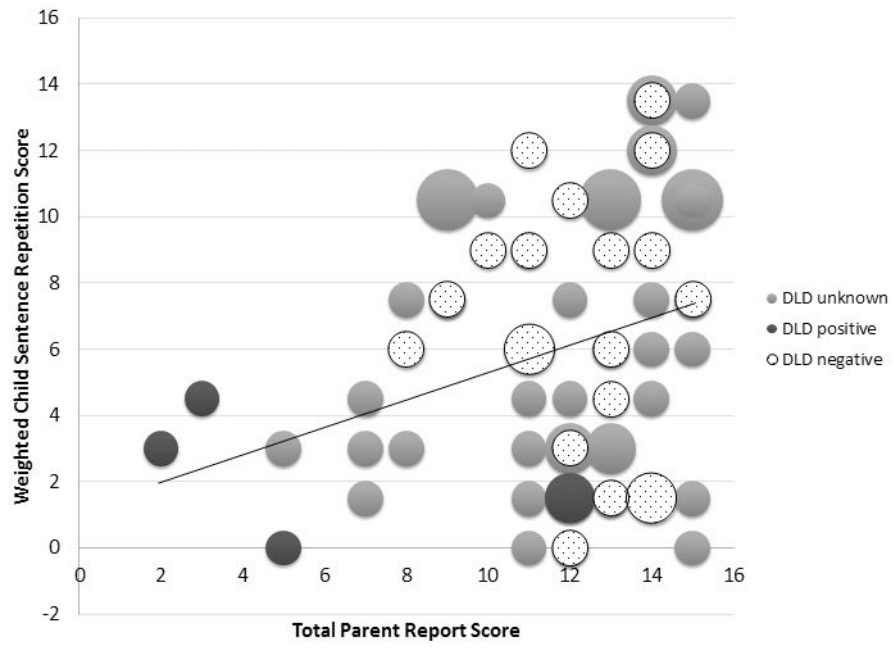


Figure 2. Relationship between parent and child scores by Developmental Language Disorder (DLD) diagnostic group. Bubble size reflects the number of participants included in each data point.

Table 1
Descriptive Statistics of Performance on the Developmental Language Disorder Screening Measure

	Full Sample		Sample Completing Follow-up Testing		Sample Not Completing Follow-up Testing	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Parent questionnaire	11.6 (3.0)	2-15	11.1 (3.4)	2-15	11.9 (2.7)	5-15
Sentence repetition task	6.0 (4.0)	0-13.5	5.8 (4.0)	0-13.5	6.1 (4.1)	0-13.5
Total Screening Score	17.6 (5.7)	5.0-28.5	16.9 (5.8)	5-27.5	18 (5.6)	8-28.5

Note. The maximum possible score for both the parent questionnaire and sentence repetition task is 15. Sentence repetition scores are reported as the number of sentences (out of 10) correctly repeated multiplied by 1.5. Total screening scores are the sum of the parent questionnaire and sentence repetition scores.

Table II

Characteristics of Groups by Diagnostic Status within the Sample that Completed Follow-Up Testing

	Developmental Language Disorder	Typical Language Development
Gender	2M, 3F	9M, 11F
Age	7.1 (1.0)	7.5 (0.6)
CELF-4 Score	71.2 (14.7)	94.2 (9.9)
ALDeQ Score	0.48 (0.20)	0.74 (0.15)
Parent questionnaire	6.8 (4.9)	12.2 (1.9)
Sentence repetition task	1.4 (1.1)	4.5 (2.6)
Total Screening Score	8.9 (4.3)	19.0 (4.2)

Note. Gender is reported as number male (M) and number female (F) within the group. All other values are reported as mean (SD). Age is reported in years. CELF-4 standard scores are reported. ALDeQ total proportion scores are reported.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table III

Cross-Tabulation of Screening Test Results by Developmental Language Disorder (DLD) Diagnostic Testing Results.

		Diagnostic Status		
		DLD	Typical Language Development	Unknown
Total Screening Score	Positive	5	1	8
	Negative	0	19	30
Parent Questionnaire Score	Positive	3	0	4
	Negative	2	20	37
Sentence Repetition Score	Positive	5	6	18
	Negative	0	14	20

Note. Results were cross-tabulated using the following cut-offs: scores of less than 14 were positive for Total Screening Score; scores less than 8 positive for parent questionnaire; scores less than 6 positive for sentence repetition score. The “unknown” column shows the distribution of scores in children who completed the screening but not the follow-up diagnostic testing.

Table IV
 Diagnostic Accuracy Estimates and Confidence Intervals for the Developmental Language Disorder Screening Test.

	Full (combined) screening tool			Parent questionnaire only			Sentence repetition only		
	Point Estimate	Lower bound, 95% CI	Upper bound, 95% CI	Point Estimate	Lower bound, 95% CI	Upper bound, 95% CI	Point Estimate	Lower bound, 95% CI	Upper bound, 95% CI
Sensitivity	1.00	1.00	1.00	0.60	0.17	1.00 ^a	1.00	1.00	1.00
Specificity	0.95	0.85	1.00 ^a	1.00	1.00	1.00	0.70	0.50	0.90
LR+	20.00	2.96	135.11	∞	N/A	∞	3.33	1.71	6.51
LR-	0.00	0.00	N/A	0.40	0.14	1.17	0.00	0.00	N/A

Note. LR+ = positive likelihood ratio; LR- = negative likelihood ratio. ∞ refers to infinity, as the positive likelihood ratio for the parent questionnaire is perfect and therefore infinite. N/A refers to values that cannot be calculated in the confidence intervals.

^aAnalyses yielded upper bounds above 1.00 for both the specificity of the combined tool (at 1.05) and the sensitivity of the parent questionnaire only (at 1.03). Because values above 1.00 are not possible for sensitivity and specificity, the upper bounds of the 95% CIs are reported as 1.00 in both cases.