Research Article

Identifying Children at Risk for Language Impairment or Dyslexia With Group-Administered Measures

Suzanne M. Adlof,^a Joanna Scoggins,^a Allison Brazendale,^a
Spencer Babb,^a and Yaacov Petscher^b

Purpose: The study aims to determine whether brief, group-administered screening measures can reliably identify second-grade children at risk for language impairment (LI) or dyslexia and to examine the degree to which parents of affected children were aware of their children's difficulties.

Method: Participants ($N = 381$) completed screening tasks and assessments of word reading, oral language, and nonverbal intelligence. Their parents completed questionnaires that inquired about reading and language development.

Results: Despite considerable overlap in the children meeting criteria for LI and dyslexia, many children exhibited problems in only one domain. The combined screening

tasks reliably identified children at risk for either LI or dyslexia (area under the curve = 0.842), but they were more accurate at identifying risk for dyslexia than LI. Parents of children with LI and/or dyslexia were frequently unaware of their children's difficulties. Parents of children with LI but good word reading skills were the least likely of all impairment groups to report concerns or prior receipt of speech, language, or reading services.

Conclusions: Group-administered screens can identify children at risk of LI and/or dyslexia with good classification accuracy and in less time than individually administered measures. More research is needed to improve the identification of children with LI who display good word reading skills.

Reading and language impairments are among the most common learning disabilities that impact children's ability to progress in school. The methods by which children with learning disabilities are identified vary within and across schools, districts, and states. However, evidence suggests that underidentification of language impairment (LI) is common, and many children with language and reading difficulties are identified late, if at all (Catts, Compton, Tomblin, & Bridges, 2012; Leach, Scarborough, & Rescorla, 2003; Lipka, Lesaux, & Seigel, 2006; Tomblin et al., 1997). This study was concerned with informing the development of methods that may improve the identification of children who have language or reading impairment that occurs alongside otherwise normal development—that is, without a known cause, such as a birth defect, vision or

hearing impairment, developmental syndrome, traumatic brain injury, or general intellectual disability.

When LI occurs in the absence of other explanatory factors (e.g., hearing impairment, intellectual disability, or other developmental syndrome), it has traditionally been termed specific language impairment (SLI; Leonard, 2014).¹ The estimated prevalence of SLI in kindergarten students is 7.4% (Tomblin et al., 1997), but children with SLI frequently go undiagnosed (Prelock, Hutchins, & Glascoe, 2008). In an epidemiologic study, Tomblin et al. (1997) reported that 70% of parents of kindergartners with SLI were unaware of their children's language weaknesses prior to formal testing. Similarly, other studies involving parent interviews have found low rates of parental knowledge of preschool (Laing, Law, Levi, & Logan, 2002)

^aUniversity of South Carolina, Columbia

^bFlorida State University, Tallahassee

Correspondence to Suzanne M. Adlof: sadlof@mailbox.sc.edu

Editor-in-Chief: Sean Redmond

Editor: Megan Dunn Davison

Received December 29, 2016

Revision received May 12, 2017

Accepted August 1, 2017

https://doi.org/10.1044/2017_JSLHR-L-16-0473

¹Recently, there has been considerable discussion about diagnostic terminology for child language disorders and a proposal to replace SLI with the term “developmental language disorder” (Bishop, 2014; Bishop, Snowling, Thompson, Greenhalgh, & CATALISE-2, 2017; Rice, 2014). In this article, we have used the label SLI to maintain an explicit link to prior studies examining the overlap between SLI and dyslexia.

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

and school-age children's LI (Conti-Ramsden, Simkin, & Pickles, 2006).

Dyslexia is defined as a specific learning disability that is characterized by difficulties with accurate and/or fluent word reading and spelling, which is unexpected in the context of other cognitive abilities and the use of effective instructional practices (Lyon, Shaywitz, & Shaywitz, 2003). Estimates of the prevalence of dyslexia vary, depending on identification criteria, but generally range between 5% and 10% (Catts, Adlof, Hogan, & Weismer, 2005; Shaywitz, Shaywitz, Fletcher, & Escobar, 1990). We are unaware of current estimates of identification rates for dyslexia. Because reading skills are directly taught and measured in schools, we hypothesize that rates of underidentification could be lower for children with dyslexia than for children with SLI.

Both SLI and dyslexia have a negative influence on the development of reading comprehension skills in affected children (Catts, Bridges, Little, & Tomblin, 2008; Snowling, Muter, & Carroll, 2007). For children with dyslexia, reading comprehension is primarily impeded by difficulty in recognizing printed words; however, some children with dyslexia also experience oral language weaknesses that can exacerbate comprehension difficulties (Catts et al., 2005; McArthur, Hogben, Edwards, Heath, & Mengler, 2000; Ransby & Swanson, 2003; Snowling, Bishop, & Stothard, 2000). Turning to children with SLI, the same deficits that make comprehension of oral language difficult (e.g., deficits in semantics, syntax, and discourse skills) also impede comprehension of written language (Botting, Simkin, & Conti-Ramsden, 2006; Kelso, Fletcher, & Lee, 2007). In addition, many children with SLI (but not all) experience word reading difficulties (Catts et al., 2005; Kelso et al., 2007).

Separate Versus Co-occurring Reading and/or Language Impairment

In the last two decades, much attention has been paid to the frequent co-occurrence of oral language and reading impairments, including SLI and dyslexia (Bishop & Snowling, 2004; Catts et al., 2005; McArthur et al., 2000; Ramus, Marshall, Rosen, & van der Lely, 2013). In a large sample drawn from an epidemiologic study, Catts et al. (2005) found that 18%–36% of children with SLI in kindergarten met criteria for dyslexia in the later school grades, and 15%–19% of children with dyslexia between second and eighth grades had a history of SLI in kindergarten. Reported rates of overlap between the two conditions have been larger in studies that employed convenience samples drawn from clinics. For example, McArthur et al. (2000) found that 55% of children with dyslexia had impaired oral language and 51% of children with SLI had a reading disability. Catts et al. (2005) hypothesized that clinical samples may be more likely to include more severely impaired individuals who are more likely to have concomitant conditions than population-based samples. That is, parents and teachers may be more likely to notice and refer for

evaluation or intervention children who have difficulties in both word reading and broader language skills than children with problems in just one domain. In a similar vein, studies of 8- to 10-year-old “poor comprehenders,” children who demonstrate poor reading comprehension despite good word recognition (and many of whom also meet criteria for SLI), have also reported that parents and teachers are frequently unaware of their children's language deficits until formal testing occurs (Adlof & Catts, 2015; Nation, Clarke, Marshall, & Durand, 2004). Nation and colleagues (2004) suggested that difficulties with speech articulation and accurate or fluent reading may be more obvious to parents and teachers, and therefore, children with these problems may be more likely to be referred to clinical specialists for evaluation than children with deficits in language but not speech or word reading.

Approaches to Language and Reading Screening

In an effort to improve the identification of children needing special education services, many schools have implemented universal screening procedures, whereby all children undergo a brief assessment to determine whether they are meeting developmental benchmarks or whether further evaluation is needed. Universal screens for speech and language difficulties are often administered at preschool or kindergarten orientations. After kindergarten, speech and language screens are generally administered in schools only in response to a teacher or parent request (Ehren & Nelson, 2005). In schools that use response to intervention (RTI) frameworks, universal screening for reading difficulties is usually conducted at the beginning of each primary grade, with subsequent benchmark assessments administered in the middle and at the end of the school year (Reschly, 2014).

In schools that use RTI frameworks for identification and service delivery, universal screens are an important component of Tier 1 (Davis, Lindo, & Compton, 2007; Reschly, 2014). In such frameworks, children who score below the cut point on a universal screen at Tier 1 are then referred to Tier 2, where they receive more intensive instruction and more frequent progress monitoring to determine whether further evaluation or intervention is needed. However, screening procedures can also be used in schools that use more traditional means of identifying children who qualify for special education services; in these cases, a failed screen typically leads to a referral for a comprehensive evaluation. In both RTI and traditional frameworks, the purpose of screening is to quickly identify individuals who may benefit from further assessment or observation. Thus, the screening measure itself should be cost and time efficient and relatively easy to implement (Jenkins & Johnson, n.d.). For the screen to be useful in practice, it must also yield high classification accuracy, such that it reliably distinguishes individuals who do or do not have poor language or reading abilities. In practice, educators are sometimes more willing to accept a higher rate of false positive errors (where children without impairment are flagged

as needing progress monitoring or assessment) than false negative errors (where children with true impairments are missed). However, it is important to note that both types of errors can lead to misallocation of resources, which can be quite costly (Davis et al., 2007).

Most existing screening tools intended for identifying children at risk for LI are designed for individual administration. Examples of commercially available screens for students in the primary and elementary school grades include the Clinical Evaluation of Language Fundamentals—Fifth Edition (Wiig, Semel, & Secord, 2013) and the Diagnostic Evaluation of Language Variation—Screening Test (DELV-ST; Seymour, Roeper, & de Villiers, 2003). Researcher-developed screening protocols have also been tested, including psycholinguistic tasks from experimental studies (e.g., nonword repetition and sentence repetition; Archibald & Joanisse, 2009). The administration time for these individually administered screens ranges from 5 to 20 min, and scoring time is generally just a few minutes. Classification accuracy statistics for these screens range from fair to good, with sensitivity estimates from 81% to 93% and specificity estimates from 57% to 87%;² however, it is not known how well these screens perform for children with co-occurring language and word reading impairments versus LIs that occur in children with typical word reading abilities.

Both individual- and group-administered screens are available to identify children at risk for reading difficulties. Examples of individually administered screens include the Predictive Assessment of Reading (Wood, 2013; Wood, Hill, Meyer, & Flowers, 2005), which requires approximately 15 min per student, as well as the widely used Dynamic Indicators of Basic Early Literacy Skills (Good & Kaminski, 2002), which requires 5 min per student per subtest. Examples of group-administered reading screens include the Test of Silent Word Reading Fluency (TOSWRF; Mather, Hammill, Allen, & Roberts, 2004) and the Test of Silent Reading Efficiency and Comprehension (Wagner, Torgesen, Rashotte, & Pearson, 2010); each can be administered in approximately 5–10 min. The reference standard used to measure classification accuracy has varied across examinations of each screen, but accuracy statistics range from fair to good for both individual- and group-administered screens, with sensitivity of 77%–93% and specificity of 65%–92% (Johnson, Pool, & Carter, 2011; Mather et al., 2004; Smolkowski & Cummings, 2015; Wagner et al., 2010; Wood et al., 2005). Similar to the case of oral language screens, it is not known how well these reading screens perform for children with co-occurring language and word reading

impairments versus reading impairments that occur in children with typical language abilities.

Study Purpose

Children with language and reading impairments experience increased risk for academic failure, underemployment, and negative socioemotional outcomes compared to their peers with good reading and language skills (Daniel et al., 2006; Johnson, Beitchman, & Brownlie, 2010; Mugnaini, Lassi, La Malfa, & Albertini, 2009). It is important to identify children with language or reading impairment as early as possible so that interventions can be implemented to facilitate positive outcomes. Previous studies have suggested that children with LI may be particularly likely to “fly under the radar” (e.g., Tomblin et al., 1997) until subsequent reading problems become apparent (Catts et al., 2012; Nation et al., 2004). The preceding review revealed several language screening tools that provide satisfactory classification accuracy, but all were individually administered. When considered for universal screening at scale—at classroom, school, and district levels—individually administered screens can be highly costly and time consuming. However, a sufficiently accurate, group-administered language screen could provide a relatively time- and cost-effective means for schools to identify children who need further observation or evaluation of their oral language skills.

This study was part of a larger project involving children with SLI, dyslexia, or typical development, which used group screenings as part of a recruitment strategy (Adlof, 2016). The screening battery included two brief, norm-referenced measures intended for group administration, including one that primarily targeted word reading skills (TOSWRF) and one that targeted sentence comprehension (Listening Comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation [GRADE LC]; Williams, 2001); administration of both screens to an entire classroom required less than 30 min. The first objective of this study was to evaluate the ability of these brief, group-administered screening measures to identify children at risk for language and/or reading impairment. Given prior studies of group-administered reading screens (Johnson et al., 2011; Mather et al., 2004) and the established rate of overlap of reading and language impairment (Catts et al., 2005; McArthur et al., 2000), we hypothesized that group-administered screens would provide acceptable levels of classification accuracy for identifying children with reading and language impairment, overall. However, no previous studies have evaluated screening accuracy for separate versus co-occurring word reading or oral language problems.

The second study objective was to examine the extent to which parents of children meeting standard criteria for LI or dyslexia would be aware of their children’s difficulties. Past studies of children with language or reading comprehension impairment (e.g., Adlof & Catts, 2015; Nation et al., 2004; Tomblin et al., 1997) led us to hypothesize that

²The DELV-ST manual does not report sensitivity and specificity, but we estimated them from the data provided in the manual. The DELV-ST provides four score classifications with regard to risk for language impairment: lowest risk, low-medium risk, medium-high risk, and highest risk. For the calculation of sensitivity and specificity, we considered the first two categories as screening passes and the latter two categories as screening failures.

few parents would have concerns about their children's language development. We are unaware of any recent studies examining parent knowledge of dyslexia in unselected samples; thus, this study was the first to compare rates of parent knowledge of children with separate versus co-occurring LI and dyslexia. We hypothesized that parents of children with combined dyslexia and LI would be more likely to report concerns about their children's skills than parents of children with difficulties in a single domain.

Method

This study was conducted as part of a larger study involving children with SLI, dyslexia, or typical development, and the screening battery was used as part of the recruitment strategy (Adlof, 2016). Study procedures were approved by the University of South Carolina Institutional Review Board. Two cohorts of children participated in academic year 2013–2014 (Year 1) and 2014–2015 (Year 2), respectively; their data were pooled for the analyses in this study. The study procedures involved two steps: (a) classroom screening and (b) individual assessment.

Classroom Screening

Trained research assistants administered language and reading screening measures to second-grade classrooms. All students in each classroom were screened at the same time.

Participants

Participants were drawn from 37 second-grade classrooms in Year 1 and 40 second-grade classrooms in Year 2. In total, the screening sample included 1,496 cases, including 732 children in Year 1 and 764 children in Year 2. The mean age of the students was 7.8 years ($SD = 0.43$). Thirty-two of the Year 1 classrooms and all of the Year 2 classrooms were from a single school district, with district demographics for second-grade students as follows: 53% boys, 47% girls; 64% White–non-Hispanic, 27% Black–non-Hispanic, 6% Hispanic, 3% two or more races, and < 1% American Indian or Asian/Pacific Islander (National Center for Education Statistics, 2015). Fifty-six percent of children in the district qualified for free or reduced lunch (South Carolina Department of Education, 2017). The remaining five classrooms in Year 1 were from a single school in a different district, with the following district demographics for second-grade students: 48% boys, 52% girls; 43% White–non-Hispanic, 30% Black–non-Hispanic, 21% Hispanic, < 1% American Indian/Alaskan, 1% Asian/Pacific Islander, 5% two or more races (National Center for Education Statistics, 2015). Fifty-eight percent of students in the school qualified for free or reduced lunch (South Carolina Department of Education, 2017). To protect student identities, students' screening forms were marked with a numeric identification code, and classroom teachers maintained the link between student names and codes.

Screening Measures

The GRADE LC (Williams, 2001) was selected to screen for potential LI. This subtest contains 17 items, intended to assess vocabulary and grammar as well as higher-level language structures (e.g., inference, idiom comprehension), and administration time is approximately 20 min. Children were given a booklet of picture sets. For each item, the examiner read a brief passage (usually a sentence) aloud, and children were asked to mark which picture, from a set of four, best matched what they heard. The GRADE manual reports coefficient alpha = .60 and split-half reliability = .75 for Level 2, Form A, as was used in this study. Although these reliabilities are less than desired, we were unaware of other group-administered, norm-referenced oral language measures for second-grade students. According to the GRADE manual, the low reliabilities may be partially related to ceiling effects among children with no oral language problems who generally find the LC subtest very easy. Thus, it was possible that this subtest could identify children with language difficulty, even if it was not reliable at discriminating average from very good performance.

Form A of the TOSWRF (Mather et al., 2004) was selected to screen for potential word reading difficulties. In the TOSWRF, children were presented with several lines of unrelated words without spaces between them and instructed to draw a line between as many real words as possible within a 3-min time limit. Students were instructed to complete each row of text sequentially. According to the manual, when a student skips more than one row, the score is considered invalid. The test manual reports that test–retest reliability is .92. Because it is a speeded test, coefficient alpha is not calculated for the TOSWRF.

Individual Assessment

Following screening, some students were invited to participate in follow-up individual testing. A planned missing data design was used in this study (Enders, 2010), such that diagnostic assessments were deliberately administered to some individuals and not others. Teachers were provided a report of scores for all students in their classroom and asked to distribute study information to children who met specific scoring criteria, which were intended to identify likely candidates for the larger project. In Year 1, we invited children whose screening scores suggested that they might have language and/or reading difficulties and children who obtained good scores on both the reading and language screening measures. Thus, teachers were asked to distribute invitations to students who met any of the following criteria: (a) scored ≤ 25 th percentile on the TOSWRF, (b) had a GRADE LC stanine score of ≤ 3 , or (c) scored ≥ 40 th percentile on TOSWRF and had a GRADE LC stanine score of > 5 . The criteria were modified for Year 2, such that more students were invited in order to meet the recruitment goals for that year. Specifically, in Year 2, the GRADE LC cutoff for “poor” performance was changed

to a stanine score of ≤ 4 , and the corresponding cutoff for “good” performance was changed to a stanine score of ≥ 5 .

Parents were provided with a packet of study information, including informed consent documents and a parent questionnaire, and invited to enroll their children in further study activities, if desired. Eligibility criteria for children to participate in follow-up testing were (a) written informed consent from the children’s parents and verbal assent from the child; (b) identification as a native English speaker; and (c) no reports of uncorrected vision, hearing, or other impairment that would interfere with the child’s ability to complete the individual assessments. The parent questionnaire was used to verify eligibility. In addition to basic demographic questions and questions about hearing, vision, motor, and other physical/medical history, parents were asked two questions that were used to assess their awareness of potential language or reading difficulties in their child. Specifically, parents were asked, “Has your child ever received speech, language, reading, or other special education services? If yes, please describe,” and “Do you have any concerns about your child’s reading or language abilities? If yes, please describe.”

Upon receipt of informed consent and parent questionnaires, follow-up tests were individually administered over one to three sessions, as needed, given student and classroom schedules. The average amount of time between screening and follow-up testing was 92 days ($SD = 28.3$ days). Follow-up testing occurred in quiet areas of the school with limited distractions. Follow-up assessments were video- and audio-recorded and scored from the recordings by trained research assistants. Gift cards for school supplies were given to each class in recognition of student and teacher time and effort in participating in the study.

Participants

A total of 1,121 students were invited to participate in individual assessments. Informed consents were returned for 485 students. The final analysis sample included 381 children who met all inclusionary criteria, as determined by the parent questionnaire, and for whom both screening measures (TOSWRF, GRADE LC) and both diagnostic measures (Clinical Evaluation of Language Fundamentals–Fourth Edition [CELF-4; Semel, Wiig, & Secord, 2004] and Woodcock Reading Mastery Tests–Third Edition [WRMT-III; Woodcock, 2011]) were available. These participants had a mean age of 7.74 years ($SD = .40$), and 54% were girls (46% boys). Their reported races were 63% White, 29.7% Black/African American, and 3.6% other/multiple races, with 3.7% not reported. Reported ethnicities were 2% Hispanic/Latino, 64% non-Hispanic/Latino, and 34% not reported.

Language Assessment

The Core Language Score from the CELF-4 (Semel, Wiig, & Secord, 2004) was used to identify children with LI. The four subtests reflected in the composite score are

Concepts and Following Directions, which measures the child’s ability to understand and follow verbal instructions of increasing length and complexity; Word Structure, which assesses morphology and pronoun use; Recalling Sentences, which assesses the child’s ability to listen to spoken sentences of increasing length and complexity and repeat them aloud verbatim; and Formulated Sentences, which assesses the child’s ability to generate a spoken sentence using a word provided by the examiner. A few children in the sample were 9 years or older. These children did not complete the Word Structure subtest, but instead completed the Word Classes subtest, which assesses a child’s ability to understand and explain semantic relationships between words. According to the test manual, the internal consistency reliability of the Core Language score of the CELF-4 ranges from .94 to .95 for the age groups represented in our study (7;0–9;11 [years;months]). In this study, children who scored at least 1 SD below the mean (standard score ≤ 85) were classified as having LI. According to the CELF-4 test manual, this cut score results in 100% sensitivity and 82% specificity of classification.

Dialectical Considerations

This study was conducted in a region of the country where many people speak a nonmainstream dialect such as African American English or Southern White English. Because the CELF-4 contains some items that test morphosyntactic features that contrast between mainstream and nonmainstream American English dialects, it was important to ensure that our language classifications were valid for individuals who spoke a nonmainstream dialect. That is, it was important to ensure that children with typical language abilities who spoke a nonmainstream dialect were not misclassified as having LI (cf. Hendricks & Adlof, 2017). We used the DELV-ST (Seymour, Roeper, & deVilliers, 2003) to verify the LI status for children who spoke a nonmainstream dialect. The DELV-ST is divided into two parts. Part I, Language Variation Status, is used to measure the amount of nonmainstream dialect feature use in the child’s spoken language productions. Raw scores from Part I are used to derive one of three classifications: strong variation, some variation, and no variation. Part II, Diagnostic Risk Status, is used to assess a child’s risk for LI, with four possible classifications of diagnostic risk: highest risk, medium-high risk, low-medium risk, and lowest risk. The DELV-ST manual reports reliability in terms of interexaminer decision consistency. According to the manual, 92% of participants in the DELV-ST reliability sample were classified the same or within one category for language variation status (Part I), and 84% of participants were classified the same or within one category for diagnostic risk status (Part II) on both screening occasions.

In the current study, most of the children who initially scored below the CELF-4 cutoff and showed “some” or “strong” variation on the DELV-ST Part I were also determined to be of medium-high to high risk for LI on the DELV-ST Part II. The Appendix contains a flowchart

detailing how dialectical variations were considered in the assignment of LI status in the current study. Following this process, 135 of 140 (96%) of the children who initially scored ≤ 85 on the CELF-4 were classified as language impaired. The remaining five children who initially scored below the CELF-4 cutoff were ultimately classified as having typical language based on adjustments for dialectical variation.

Word Reading Assessment

The Word Identification and Word Attack subtests from the WRMT-III (Woodcock, 2011) were used to identify children with word reading difficulties commensurate with dyslexia. The Word Identification subtest requires children to read real English words of increasing difficulty, whereas the Word Attack subtest requires children to read pronounceable pseudowords of increasing difficulty. According to the test manual, split-half reliability for the Word Identification subtest is .94 and split-half reliability for the Word Attack subtest is .92 for second-grade students. These subtests combine to form a cluster score, Basic Skills. According to the test manual, reliability for the Basic Skills cluster is .96 for second-grade students, using a formula recommended by Feldt and Brennan (1989) and Guilford (1954). Children who scored at least 1 *SD* below the mean on the Basic Skills cluster were classified as dyslexic. Although the WRMT-III manual does not provide information regarding sensitivity and specificity, this cutoff is comparable to other studies that have used the WRMT-III (or previous versions) for identifying dyslexia (e.g., Catts et al., 2005; Joanisse, Manis, Keating, & Seidenberg, 2000; Siegel, 2008). The WRMT-III manual provides descriptive statistics for a sample of individuals with learning disabilities in the area of reading. Our selection of an 85 standard score cutoff would identify the majority of individuals in this subgroup, who achieved a mean of 76.8 (*SD* = 11.9).

Nonverbal Intelligence

The Test of Nonverbal Intelligence–Fourth Edition (TONI-4; Brown, Sherbenou, & Johnsen, 2010) was used to assess nonverbal cognitive skills. According to the test manual, the internal consistency of the TONI-4 is .94–.96 for the age groups represented in this study. Some participants ($n = 40$) were missing data on the TONI-4. Most of these ($n = 34$) were children who were considered typically developing (TD) based on the CELF-4 and WRMT-III but who did not participate in the larger study and thus were not administered the TONI-4. In this study, we used the TONI-4 for purely descriptive purposes, and no decisions for participant inclusion or exclusion were made on the basis of this test. Our rationale was that it was important for the screening measures to accurately classify all children at risk for language or reading impairment, regardless of whether the difficulties were “specific” to language and/or reading or commensurate with nonverbal cognitive abilities (cf. Fletcher et al., 1994; Rice, 2016). However, the vast majority of the participants in the

impairment groups showed normal nonverbal cognitive skills.³

Scoring Reliability

In Year 1, all screening measures and assessments were double scored by trained research assistants to ensure reliability, and disagreements were reconciled through discussion with the lead assessor and/or principal investigator (third author and first author, respectively). In Year 2, all scorers were required to pass a scoring test before scoring protocols independently. Scorers kept a log of tests that they had scored, and a random sample of at least 20% of each scorer’s list was double-scored for assessing reliability. Because participants wrote their own answers on screening protocols and the initial scorers marked up the protocols, screening reliability scorers were not blind to initial scores. Reliability was assessed as the percentage of agreements on summary scores and was 94% for TOSWRF standard score and 99% for GRADE LC stanine score. Reliability scorers for all individually administered assessments used blank protocols and video/audio recordings of the assessments and were blind to initial scores. Reliability was assessed as the by-item agreement for each of the individually administered measures and was 92.8% for CELF-4, 96.5% for WRMT-III, 99.3% for TONI-4, and 92% for DELV-ST.

Results

Descriptive statistics for the screening measures are provided in Table 1 for the full screening sample and for the subset of children who participated in individual testing. Note that, within the full screening sample, some children only completed one measure and/or had invalid scores for the TOSWRF, leading to unequal *N*s for these two measures. The mean scores of the screening sample and the individual testing sample were not significantly different for either the TOSWRF ($t = 1.68, p = .09$) or the GRADE LC ($t = 0.82, p = .41$). In the individual testing sample, LI and dyslexia (DYS) criteria were applied independently of each other. Thus, each child could meet the criteria for one (either LI-only or DYS-only), both (LI+DYS), or no impairment (TD). In this sample, 54% of children meeting the criteria for LI also met the criteria for DYS, and 71% of children meeting the criteria for DYS also met the criteria for LI (see Figure 1). Thus, there was considerable

³Specifically, 89% of the participants classified as LI (120/135) had TONI-4 standard scores of > 85 , 7% (10/135) had TONI-4 standard scores of between 75 and 85, one had a TONI-4 standard score of 68, and 3% (4/135) were missing TONI-4 data. Likewise, 88% of the participants meeting the dyslexia criteria (91/103) had TONI-4 standard scores of > 85 , 7% (7/103) had TONI-4 standard scores between 75 and 85, one had a TONI-4 standard score of 68, and 4% (4/103) were missing TONI-4 data. Within the group of participants classified as having neither LI nor dyslexia (i.e., typically developing group), 84% (181/216) had TONI-4 scores of > 85 , one had a TONI-4 standard score of 79, and 16% (34/216) were missing TONI-4 data.

Table 1. Means and standard deviations for screening measures in screening sample and diagnostic sample.

Measure	Screening sample			Diagnostic sample		
	N	M	SD	N	M	SD
TOSWRF SS	1,429	100.14	13.88	381	98.79	14.04
GRADE LC stanine	1,440	4.76	2.11	381	4.66	2.16

Note. TOSWRF SS = Test of Silent Word Reading Fluency Standard Score; GRADE LC = Group Reading Assessment and Diagnostic Evaluation Listening Comprehension subtest.

overlap between word reading and language difficulties, but there were also many children who showed difficulty in one domain and typical performance in the other. Table 2 provides descriptive statistics for children who were classified as LI-only, DYS-only, LI+DYS, or TD.

Screening Performance for LI and/or DYS

Our first research objective was to determine how well brief classroom screens could identify second-grade students with LI and/or DYS, as determined by individually administered diagnostic measures. We began by examining correlations between group-administered and individually administered measures. A small but significant correlation between scores on the two group screening measures (TOSWRF and GRADE LC) was observed in both the full screening sample $r = .21$ ($p < .001$) and the smaller individualized testing sample $r = .24$ ($p < .001$). Table 3 displays correlations between scores on the group screening measures and scores on the diagnostic measures (WRMT-III and CELF-4) for the individual assessment sample. Although both group screens were significantly correlated with both diagnostic assessments, we were surprised that the correlation between the word reading screening measure (TOSWRF) and the language diagnostic measure

Figure 1. Venn diagram illustrating the overlap between students meeting criteria for language impairment (LI) and/or dyslexia (DYS). Of the 135 children meeting the criteria for LI, 73 (54%) also met the criteria for DYS. Of the 103 children meeting the criteria for DYS, 73 (71%) also met the criteria for LI. LI-only = language impairment without dyslexia; DYS-only = dyslexia without language impairment; LI+DYS = combined language impairment and dyslexia.

(CELF-4; $r = .58$, $p < .001$) was larger than the correlation between the language screening (GRADE LC) and language diagnostic measure (CELF-4; $r = .38$, $p < .001$).

Logistic regression and receiver operating characteristic (ROC) curves were used to examine how well the screening measures differentiated children with LI or DYS from children with typical language and reading skills. Recall that participant recruitment procedures resulted in some data being missing completely at random (MCAR). Little's test of MCAR was used to statistically evaluate the extent to which the data aligned with this assumption of the data (Little, 1988). Results for the MCAR test affirmed that an assumption of MCAR was tenable, $\chi^2(2) = 2.77$, $p = .251$, indicating that the missing data could be addressed via estimation methods. Therefore, we used maximum likelihood estimation in the logistic regression models as a means to leverage the available data. The logistic procedure in SAS (SAS Institute Inc., 2013) was used for the logistic regression and area under the curve (AUC) estimation (Schatschneider, 2013). Note that in each model the N for the typical comparison group varied for each prediction as children in the LI-only group (who met criteria for LI but not DYS) were considered to have typical reading skills and children in the DYS-only group (who met criteria for DYS but not LI) were considered to have typical language skills.

We ran three groups of logistic regression models predicting the risk of LI (including LI-only and LI+DYS) versus typical language (TD and DYS), DYS (including DYS-only and LI+DYS) versus typical reading (TD and LI), and either LI or DYS (including LI-only, DYS-only, and LI+DYS) versus TD. For the prediction of risk for LI versus typical language, GRADE LC scores were entered in the first logistic regression model; the second model included both GRADE LC and TOSWRF scores. In the prediction of risk for DYS versus typical reading, the modeling procedure was reversed. In the prediction of either impairment, a single model was run with both predictors. The predicted probabilities from the models containing both predictors were saved as composite risk scores for use in ROC curve analysis.

ROC curves were used to evaluate overall classification accuracy. A ROC curve is a plot of the sensitivity of the screening measure (i.e., the true positive rate) against its false positive rate (i.e., $1 - \text{specificity}$) for every possible screening score. The AUC provides an estimate of the screening accuracy: For all possible pairs of individuals, where one member of the pair is impaired and the other is not impaired, the AUC indicates the percentage of times the screen will assign a higher risk status to the impaired member of the pair. A measure with chance-level accuracy will result in a diagonal line and an AUC value of .50. As a rule of thumb, AUC values between .7 and .8 are usually considered acceptable, values between .8 and .9 are considered excellent, and values above .9 are outstanding (Hosmer & Lemeshow, 2000). For each risk outcome (LI, DYS, either impairment), we generated ROC curves for each screening measure individually, as well as the composite risk score derived from the logistic regression models (i.e., predicted probability score from full model).

Table 2. Subgroup descriptive statistics for screening and diagnostic measures and parent report.

Measure	TD	LI-only	DYS-only	LI+DYS
	(<i>n</i> = 216)	(<i>n</i> = 62)	(<i>n</i> = 30)	(<i>n</i> = 73)
TOSWRF SS mean (<i>SD</i>)	105.62 _a (11.66)	95.66 _b (12.14)	88.23 _c (7.77)	85.56 _c (10.65)
GRADE LC stanine mean (<i>SD</i>)	5.24 _a (2.02)	3.58 _b (1.96)	4.93 _a (2.12)	3.78 _b (2.10)
CELF-4 Core Language SS mean (<i>SD</i>)	102.47 _a (9.55)	78.13 _b (6.59)	94.80 _c (6.96)	73.04 _d (8.87)
WRMT-III Basic Skills SS mean (<i>SD</i>)	104.63 _a (11.14)	94.37 _b (7.73)	80.23 _c (3.32)	74.47 _d (7.38)
TONI-4 SS ^a mean (<i>SD</i>)	106.78 _a (8.92)	100.42 _b (9.49)	99.86 _b (8.46)	95.09 _c (8.77)
Parent concern about reading or language skills	7.4% _a	24.2% _b	46.7% _c	37.0% _{b,c}
Parent-reported child history of speech, language, or reading service	6.9% _a	14.5% _{a,b}	20.0% _{b,c}	38.4% _c
Combined parent concern and/or history of service	13.9% _a	29.0% _b	60.0% _c	56.2% _c

Note. Groups that share subscripts (a–d) are not significantly different from each other, $p > .05$. All other comparisons are significantly different, $p < .05$. TD = typical development; LI-only = language impairment without dyslexia; DYS-only = dyslexia without language impairment; LI+DYS = both language impairment and dyslexia; TOSWRF = Test of Silent Word Reading Fluency; SS = standard score; GRADE LC = Group Reading Assessment and Diagnostic Evaluation Listening Comprehension subtest; CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition; WRMT-III = Woodcock Reading Mastery Tests–Third Edition; TONI-4 = Test of Nonverbal Intelligence–Fourth Edition.

^aThe sample size for the TONI-4 was smaller than other assessments (TD $n = 182$, LI-only $n = 60$, DYS-only $n = 28$, LI+DYS $n = 71$).

Tables 4–6 report the results of the logistic regression models for each predicted outcome, and Table 7 reports the results of the ROC curve analyses. For the prediction of risk of LI, the model including GRADE alone was significantly better than chance, resulting in a nearly acceptable level of classification accuracy (AUC = .699). However, the full model including TOSWRF was significantly better, resulting in greatly improved classification accuracy (AUC = .792). A different pattern of results was found for the prediction of risk of DYS, such that TOSWRF scores alone resulted in a very high level of classification accuracy (AUC = .857), and the addition of GRADE scores did not improve the prediction. For the prediction of risk for either impairment, both TOSWRF and GRADE LC scores were statistically significant in the model. The AUC for the composite risk score (.842) was considerably greater than the AUC for the GRADE LC alone (.676) and

only slightly larger than the AUC for the TOSWRF alone (.833).

Overall, the results indicated that the addition of GRADE LC in the combined model made a small improvement to overall classification accuracy over the model with TOSWRF by itself. To examine how well the screens identified children with single versus combined deficits, we examined for each subgroup (i.e., LI-only, LI+DYS, DYS-only, TD) the distribution of TOSWRF scores and composite risk scores (i.e., predicted probability) from the final model predicting risk for either LI or DYS. These two measures showed comparable AUC values; GRADE LC was not considered by itself due to its substantially lower AUC. The boxplots in Figures 2 and 3 show the range of scores exhibited by each subgroup on the composite risk score and the TOSWRF. The dotted line plotted on each graph displays the cut score associated with an overall sensitivity of .8. As

Table 3. Correlations between group screening and individual assessment measures in the individual assessment sample ($N = 381$).

Measures	1	2	3	4
1. TOSWRF SS	1			
2. GRADE LC stanine	.24*	1		
3. CELF-4 Core Language SS	.58*	.38*	1	
4. WRMT-III Basic Skills SS	.71*	.23*	.69*	1

Note. TOSWRF = Test of Silent Word Reading Fluency; SS = standard score; GRADE LC = Group Reading Assessment and Diagnostic Evaluation Listening Comprehension subtest; CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition; WRMT-III = Woodcock Reading Mastery Tests–Third Edition.

* $p < .001$

Table 4. Results of logistic regression models predicting risk of language impairment ($n = 135$) versus typical language ($n = 246$).

Model	Predictor	β	SE β	Wald's χ^2	df	<i>p</i>	OR
1 _a	GRADE LC	−0.36	0.06	37.32	1	< .001	0.70
	Constant	1.06	0.28	14.95	1	< .001	
2 _b	GRADE LC	−0.31	0.07	22.36	1	< .001	0.74
	TOSWRF	−0.08	0.01	54.99	1	< .001	0.93
	Constant	8.25	1.04	63.36	1	< .001	

Note. Subscript a: Model 1 summary statistics: $\chi^2 = 44.14$, $df = 1$, $p < .001$; −2 log likelihood = 456.95. Subscript b: Model 2 summary statistics: $\chi^2 = 113.48$, $df = 2$, $p < .001$; −2 log likelihood = 387.60. GRADE LC = Group Reading Assessment and Diagnostic Evaluation Listening Comprehension subtest; TOSWRF = Test of Silent Word Reading Fluency.

Table 5. Results of logistic regression models predicting risk of dyslexia ($n = 103$) versus typical reading ($n = 278$).

Model	Predictor	β	SE β	Wald's χ^2	df	p	OR
1 _a	TOSWRF	-0.12	0.01	82.47	1	< .001	0.89
	Constant	10.02	1.18	72.18	1	< .001	
2 _b	TOSWRF	-0.11	0.01	78.08	1	< .001	0.89
	GRADE LC	-0.05	0.07	0.48	1	.483	0.96
	Constant	10.08	1.19	72.21	1	< .001	

Note. Subscript a: Model 1 summary statistics: $\chi^2 = 82.47$, $df = 1$, $p < .001$; $-2 \log$ likelihood = 132.01. Subscript b: Model 2 summary statistics: $\chi^2 = 82.38$, $df = 2$, $p < .001$; $-2 \log$ likelihood = 132.49. TOSWRF = Test of Silent Word Reading Fluency; GRADE LC = Group Reading Assessment and Diagnostic Evaluation Listening Comprehension subtest.

can be inferred from the boxplots, the three impaired groups scored significantly differently than the TD group on both screening measures (all $ps < .001$). However, the LI-only group displayed a wider range of TOSWRF scores and composite risk scores than the DYS-only or LI+DYS group, such that more of the LI-only group's scores overlapped with the TD group. Thus, the screening measures were somewhat less effective at identifying LI-only participants than LI+DYS or DYS participants. Table 8 indicates the cut score, overall false positive rate, and rate of misses for the three impaired subgroups associated with overall sensitivity rates of .7, .8, and .9 for the TOSWRF and combined screen. These values suggest that the addition of the GRADE LC to the TOSWRF in the combined screen leads to small decreases in the overall false positive rate and the rate of misses in the LI-only group and a small increase in misses for the LI+DYS and DYS-only groups.

Parent Awareness of LI and/or DYS

Our second research objective was to examine whether parents of children with single deficits—either LI-only or DYS-only—were aware of their children's difficulties at the same rate as parents of children with combined LI+DYS. The last three rows in Table 2 indicate the percentage of parents of children in each group who (a) reported concerns about their child's language or reading skills; (b) reported that their child had previously received services for speech,

Table 6. Results of logistic regression models predicting risk of either language or reading impairment ($n = 165$) versus typical development ($n = 216$).

Predictor	β	SE β	Wald's χ^2	df	p	OR
TOSWRF	-0.10	0.01	79.17	1	< .001	0.90
GRADE LC	-0.25	0.06	15.38	1	< .001	0.78
Constant	10.99	1.18	87.55	1	< .001	

Note. Model summary statistics: $\chi^2 = 152.56$, $df = 2$, $p < .001$; $-2 \log$ likelihood = 368.77. TOSWRF = Test of Silent Word Reading Fluency; GRADE LC = Group Reading Assessment and Diagnostic Evaluation Listening Comprehension subtest.

language, or reading; or (c) reported either concerns or prior services. Z tests (two-tailed) of the collapsed responses indicated that parents of children in the DYS-only and LI+DYS groups reported concerns and/or prior receipt of services significantly more often than parents of children in the LI-only group (LI-only vs. DYS-only, $p = .004$; LI-only vs. LI+DYS, $p = .001$), whereas the difference between the LI+DYS group and DYS-only group was not significant ($p = .73$). Parents of children in all three impairment groups were significantly more likely to report concerns than parents of children in the TD group (all $ps < .007$). Overall, questionnaire responses suggested that a considerable percentage of parents across all three impairment groups were unaware of their children's difficulties and that parents were more likely to be aware of reading difficulties than oral language difficulties.

Discussion

This study first examined the ability of brief, classroom-administered screening measures to accurately identify second-grade students with language or reading impairment. In light of concerns about the amount of time that is spent on assessments in schools (cf. Plank & Condliffe, 2013), screening tools that provide quick identification of students who may need a closer look while protecting instructional time can be very valuable. Overall, the classification accuracies achieved by the combination of the two group-administered measures in this study were all above .8, which is considered outstanding (Hosmer & Lemeshow, 2000). These classification accuracies were as good as or better than the individually administered measures described in the introduction, which require considerably more time to administer and score. We were able to administer the combined screening battery in less than 30 min per classroom. Scoring time averaged less than 5 min per student. Thus, in total, risk status could be determined for a classroom of 20 students in approximately 2 hr, including the administration and scoring of both measures. Furthermore, no special equipment was required, as both measures required only paper and pencil. Thus, these results indicate the validity of brief, group-administered measures for identifying children at risk for language or reading impairment. Furthermore, there is potential for such group-administered tools to provide accurate identification with considerable savings compared to the time, personnel, and equipment required for individual administration of screening measures.

Overall, the results supported the usefulness of the TOSWRF to a much greater extent than the GRADE LC. The TOSWRF provided excellent classification accuracy for identifying children with DYS or LI + DYS and acceptable classification accuracy for identifying children with LI but good word reading scores. By itself, the GRADE LC was not sufficiently discriminating, but it did help improve the accuracy of classification of children with LI but good word reading scores and reduce the false positive rate (the rate of falsely classifying TD children as being at risk). The

Table 7. Areas under the curve (AUC) and 95% confidence intervals (CI) for predicting risk of language impairment (LI), dyslexia (DYS), or either impairment from individual and combined screening measures.

Prediction outcome	Impairment <i>n</i>	Typical <i>n</i>	AUC	95% CI
Risk of LI vs. typical language				
GRADE LC	135	246	.699	.644–.755
TOSWRF			.778	.730–.826
Composite			.792	.746–.838
Risk of DYS vs. typical word reading				
TOSWRF	103	278	.857	.818–.896
GRADE LC			.602	.537–.666
Composite			.852	.812–.891
Risk of LI or DYS vs. typical word reading and language				
TOSWRF	165	216	.833	.792–.875
GRADE LC			.676	.622–.731
Composite			.842	.802–.882

Note. GRADE LC = Group Reading Assessment and Diagnostic Evaluation Listening Comprehension subtest; TOSWRF = Test of Silent Word Reading Fluency.

superior discrimination of the TOSWRF for identifying word reading problems as compared to the GRADE LC's identification of language problems may be partially explained by the fact that the TOSWRF was more similar to the diagnostic reading task (WRMT-III) than the GRADE LC was to the diagnostic language task (CELF-4). Both the TOSWRF and the WRMT-III focused on the reading of single words (or pseudowords). In contrast, a wider array of tasks (following multistep directions, understanding and producing grammatical morphemes in isolation and in

sentences, repeating sentences, generating sentences, and completing morphosyntactic cloze tasks) were evaluated in the CELF-4 than in the GRADE LC.

The TOSWRF also provided better discrimination than the GRADE LC for identifying language problems. One possible explanation is that the TOSWRF likely tapped into language skills including vocabulary knowledge and metalinguistic awareness, as children were required to identify the boundaries of real words amidst other real words. Another potential explanation relates to the scaling of the

Figure 2. Distribution of scores on composite screen by subgroup. Dotted line indicates cut score associated with 80% sensitivity overall. TD = typical development; LI-only = language impairment without dyslexia; DYS-only = dyslexia without language impairment; LI+DYS = combined language impairment and dyslexia.

Figure 3. Distribution of scores on Test of Silent Word Reading Fluency by subgroup. Dotted line indicates cut score associated with 80% sensitivity overall. TD = typical development; LI-only = language impairment without dyslexia; DYS-only = dyslexia without language impairment; LI+DYS = combined language impairment and dyslexia; TOSWRF SS = Test of Silent Word Reading Fluency standard score.

GRADE LC items. Specifically, the difference in raw scores between a stanine score of 3 (11th to 23rd percentiles) and a stanine score of 6 (77th to 89th percentiles) was only three points out of a maximum of 17. As noted previously, the GRADE LC manual indicated that ceiling effects in the norming sample were related to the fact that the task is very easy for students without receptive oral language problems. However, our data suggested that the task was also relatively easy for many of the students who met the criteria for LI. Specifically, 21% of the participants meeting the criteria for LI (including 15/62 in the LI-only group and 13/73 in the LI+DYS group) missed no more than one item on the GRADE LC, and an additional 27% (including 12/62 in the LI-only group and 24/73 in the LI+DYS group) missed no more than three items. In comparison, 34.5% of children

with typical language skills (including 8/30 in the DYS-only group and 77/216 in the TD group) missed three or more items on the GRADE LC. In summary, although significant mean differences were observed between the groups with typical language skills versus LI, the distributions of both raw and stanine scores between groups showed considerable overlap. We note that the GRADE LC was used because we were unaware of other group-administered, norm-referenced oral language measures that could be administered in the same or less time for our larger project. However, it is possible that better classification accuracy could be achieved using a sentence comprehension task that involved a better selection of items or by adapting other types of tasks for group administration. Overall, our results underscore the potential of group-administered assessments

Table 8. Screening performance for language impairment + dyslexia subgroups with Test of Silent Word Reading Fluency (TOSWRF) versus combined screen.

Screening measure	Cut score	Overall sensitivity	Overall false positive rate	LI-only		DYS-only		LI+DYS	
				Numer. missed	Proportion	Numer. missed	Proportion	Numer. missed	Proportion
TOSWRF	106	.90	.49	15	.24	1	.03	1	.01
	100	.80	.27	24	.39	3	.10	6	.08
	96	.70	.17	29	.47	7	.23	13	.18
Combined screen	0.27	.90	.38	11	.18	2	.07	3	.04
	0.40	.80	.24	21	.34	4	.13	7	.10
	0.54	.70	.17	28	.45	7	.23	13	.18

Note. LI-only = language impairment without dyslexia; DYS-only = dyslexia without language impairment; LI+DYS = both language impairment and dyslexia.

for the identification of at-risk children, but there is room for improvement.

An important strength of this study was its consideration of separate versus co-occurring reading and language problems. The measures we employed resulted in accurate classification of children at risk overall, but they tended to miss children from the LI-only group at higher rates than the DYS and LI+DYS groups. We are unaware of any prior studies that have evaluated screening accuracy for separate versus co-occurring LI and DYS. However, an important question is how well children with LI, but not DYS, can be reliably identified with other available screening measures, whether group or individually administered. Children with LI are at risk for significant reading comprehension problems even if their word reading abilities appear adequate (Catts et al., 2006; Clair, Durkin, Conti-Ramsden, & Pickles, 2010; Nation et al., 2004; see also Adlof, Catts, & Lee, 2010; Spencer, Quinn, & Wagner, 2014). Thus, it is critical to have reliable methods of identifying this group of children available as early as possible so that intervention can be provided to prevent or ameliorate reading comprehension difficulties.

Our second research objective was to examine the extent to which parents of second-grade students meeting traditional criteria for language or word reading impairment might be aware of their children's difficulties, as measured by responses to a parent questionnaire indicating concerns about their children's language or reading abilities or their children's prior receipt of speech, language, or reading services. Overall, the results suggested that many to most parents of children meeting traditional criteria for language or reading impairment (40%–71%) were unaware of their children's difficulties. Of the three impairment groups, parents of children in the LI-only group (who had relatively good word reading skills) were the least likely to report concerns and/or prior receipt of services. They were significantly less likely to report concerns and/or prior history than parents of children in the DYS group and the LI+DYS group (29% vs. 60% and 56.2%, respectively). We had hypothesized that parents would be more likely to report concerns for children with both word reading and oral language problems than children with problems in only one domain. Instead, these results suggest that parents were more aware of word reading problems than oral language problems. These results converge with the findings of other studies of preschool and school-age children with SLI, in which few parents were aware of their children's language difficulties (Conti-Ramsden et al., 2006; Laing et al., 2002; Tomblin et al., 1997). Many children with LI in the early grades go on to have subsequent difficulty with reading comprehension, even if their word reading skills develop normally (Catts, Adlof, & Weismer, 2006; see also Catts et al., 2012), and studies of poor comprehenders (many of whom meet the criteria for LI) have also reported low levels of parent and/or teacher knowledge of oral language weakness (Adlof & Catts, 2015; Nation, et al., 2004). If parents are generally unaware of their children's difficulties, they may also be unlikely to seek early intervention

services, unless the risk for potential problems is brought to their attention, for example, through universal screening procedures. Our results indicated that children meeting the criteria for LI but not DYS were least likely to be identified by parents or screening measures as being at risk. Thus, more research is needed to develop efficient yet accurate screens that can identify children who present with LI in spite of good word reading ability.

To our knowledge, this study is the first since Catts et al. (2005) to examine all four groups (LI-only, DYS-only, LI+DYS, TD) within a community-based sample versus preidentified clinical samples. Although it was not a specific research question for this study, we observed that the rate of overlap between LI and DYS in the current sample was higher than expected based on Catts and colleagues' (2005) population-based sample. Catts et al. found that 18%–33% of children with a history of SLI in kindergarten went on to have DYS in second grade and 15%–19% of children with DYS in second grade had a history of SLI in kindergarten. In contrast, the current study found that 54% of children meeting the criteria for LI also met the criteria for DYS and 71% of the children with DYS also met the criteria for LI. Differences in rates of overlap among past studies have generally been attributed to differences in recruitment—clinical referral versus population-based samples—but the similarity of recruitment techniques between the two studies (see Tomblin et al., 1997, for detailed explanation of the participant screening and recruitment) suggests that other factors must be at play. It is possible that slight differences in the measures used to identify LI and DYS in each study could explain some differences. Another important difference between this study and Catts et al. (2005) is that, in the current study, the LI and DYS labels were applied based on tests administered at the same point in time, whereas SLI in Catts et al.'s study was determined by kindergarten language performance and dyslexia by second-grade reading performance. Although LI is generally thought to be a persistent condition, it is likely that at least some children's language status would have changed between kindergarten and second grade (Tomblin, Zhang, Buckwalter, & O'Brien, 2003). In addition, the participants in this study were recruited from a different geographic region and were more racially diverse than in the Catts et al. (2005) study (see also Tomblin et al., 1997). Furthermore, there may be differences between the two samples in home and school language and literacy experiences that explain differences in rates of overlap. For example, the children in the Catts et al. (2005) study were enrolled in primary school in the mid-1990s. Since that time, evidence-based reading instruction has received considerable attention in research and educational policy (Gersten et al., 2008; No Child Left Behind Act of 2001). Thus, differences between the samples in the number of children who showed poor word reading performance due to poor instruction, as opposed to genetic or neurobiologically based impairment (cf. Vellutino, Fletcher, Snowling, & Scanlon, 2004), might explain some differences in rates of overlap between this study and Catts et al. (2005). Although our

data do not allow us to draw strong conclusions, they do highlight the need for additional population-based studies to investigate separate versus co-occurring LI and dyslexia. Despite the higher rate of overlap in this study, it is important to note that some children did show unique profiles of deficit in one domain and relative strength in the other. Attending to the differences between the LI and DYS profiles is important, as they have different needs for intervention.

Before concluding, we acknowledge several caveats and limitations that should be addressed in future studies. First, this study involved a single grade level, second grade. This time point was selected because it is a time when word reading ability can be reliably assessed—after a child has had at least 2 years of reading instruction in school. However, the screening measures employed in this study may perform differently with other age groups, and they may not work well for very young children. Second, although our criteria for DYS were similar to those employed in past studies (Catts et al., 2005; Joanisse et al., 2000; Siegel, 2008), we cannot rule out the possibility that some children classified as dyslexic in this study may have had poor reading instruction. Our rationale was that all children who showed significant word reading difficulties should be identified for further observation, whether that be a comprehensive evaluation or a high-quality Tier 2 instruction, which is intended to rule out poor reading instruction as a causal factor (Reschly, 2014). However, the extent to which each type of word reading difficulty (i.e., defined with/without consideration of response to instruction) overlaps with LI is an important question for future study. Third, our conclusions about parental awareness of children's skills were drawn from responses to two open-ended questionnaire items. Although our results converge with other cited studies that have reported low parental awareness of language deficits (e.g., Adlof & Catts, 2015; Conti-Ramsden et al., 2006; Laing et al., 2002; Nation et al., 2004; Tomblin et al., 1997), it is possible that different results would be obtained with other methods of assessing parent awareness. We also did not query teachers about their awareness of students' reading or language weaknesses, but it is likely they would have a higher level of awareness than parents, especially with regard to word reading skills. Future studies could investigate whether including information obtained from teachers improves the performance of group-administered screens (cf. Whitworth, Davies, Stokes, & Blain, 1993). Fourth, although our sample size was relatively large compared to past evaluations of LI screens and of LI and DYS subgroups, it came from a single region of the country and primarily a single school district. Whereas the overall patterns observed in this study are likely valid for other geographic areas, the specific cut scores that best discriminate children at risk for LI or DYS in this sample may differ from those for other regions. Finally, although our data met statistical assumptions for data MCAR, missing data nonetheless can introduce bias regardless of mechanism or level. Thus, future studies with more complete data would be valuable to test whether the findings replicate what has been presented here.

Conclusion

This study found that brief, group-administered screening measures were able to efficiently identify at-risk children with overall high classification accuracy. The group screens provided excellent identification of children with dyslexia and co-occurring dyslexia and LI, but they were less accurate at identifying children with LI but good reading skills. In addition, we found that parents of second-grade children meeting standard criteria for LI or dyslexia were frequently unaware of their children's language and reading difficulties, and parents of children with LI but good word reading skills were least likely to report concerns. Further study is needed to consider other age groups and geographic regions and to develop protocols to reliably identify children with LI but good word reading skills.

Acknowledgments

This research was supported, in part, by funding from the National Institutes of Health (R03DC013399) to the University of South Carolina (PI: Adlof). We thank the participants of this study and the teachers and schools who assisted us with screening and recruitment and provided space and time for assessment. We thank research assistants from the SCROLL Lab at the University of South Carolina for their help with data collection and processing, including Sheida Abdi, Pooja Adarkar, Ellen Ashley, Faith Baumann, Alex Cattano, Rebecca Duross, Madison Goehring, Amanda Harris, Katie Harrison, Alison Hendricks, Alyssa Ives, Hannah Kinkead, Pooja Malhorta, Sara Mallon, Elaine Miller, Hannah Patten, Caroline Smith, Sheneka White, and Kimberly Wood. We are also grateful to Alison Hendricks for comments on an earlier version of this manuscript.

References

- Adlof, S. M. (2016, July). *Spoken word learning in children with SLI, dyslexia, and typical development*. Paper presented at the meeting of the Society for the Scientific Study of Reading, Porto, Portugal.
- Adlof, S. M., & Catts, H. W. (2015). Morphosyntax in poor comprehenders. *Reading and Writing*, 28, 1051–1070.
- Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, 43, 332–345.
- Archibald, L. M., & Joanisse, M. F. (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, 899–914.
- Bishop, D. V. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language & Communication Disorders*, 49(4), 381–415.
- Bishop, D. V., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*, 130, 858–886.
- Bishop, D. V., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2. (2017). Phase 2 of CATALISE: A multi-national and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, 58, 1068–1080. <https://doi.org/10.1111/jcpp.12721>

- Botting, N., Simkin, Z., & Conti-Ramsden, G. (2006). Associated reading skills in children with a history of specific language impairment (SLI). *Reading and Writing, 19*, 77–98.
- Brown, L., Sherbenou, R., & Johnsen, S. (2010). *Test of Nonverbal Intelligence—Fourth Edition*. Austin, TX: Pro-Ed.
- Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research, 48*, 1378–1396.
- Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278–293.
- Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research, 51*, 1569–1579.
- Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology, 104*, 166–181.
- Clair, M. C., Durkin, K., Conti-Ramsden, G., & Pickles, A. (2010). Growth of reading skills in children with a history of specific language impairment: The role of autistic symptomatology and language-related abilities. *British Journal of Developmental Psychology, 28*, 109–131.
- Conti-Ramsden, G., Simkin, Z., & Pickles, A. (2006). Estimating familial loading in SLI: A comparison of direct assessment versus parental interview. *Journal of Speech, Language, and Hearing Research, 49*, 88–101.
- Daniel, S. S., Walsh, A. K., Goldston, D. B., Arnold, E. M., Reboussin, B. A., & Wood, F. B. (2006). Suicidality, school dropout, and reading problems among adolescents. *Journal of Learning Disabilities, 39*, 507–514.
- Davis, G. N., Lindo, E. J., & Compton, D. L. (2007). Children at risk for reading failure; constructing an early screening measure. *Teaching Exceptional Children, 39*, 32–37.
- Ehren, B. J., & Nelson, N. W. (2005). The responsiveness to intervention approach and language impairment. *Topics in Language Disorders, 25*, 120–131.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.
- Feldt, L. S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 104–146). New York, NY: Macmillan.
- Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., . . . Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*, 6–23.
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/wwc/publications/practiceguides/>
- Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills—Sixth Edition*. Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from <http://dibels.uoregon.e>
- Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York, NY: McGraw-Hill Education.
- Hendricks, A. E., & Adlof, S. M. (2017). Language assessment with children who speak nonmainstream dialects: Examining the effects of scoring modifications in norm-referenced assessment. *Language, Speech, and Hearing Services in Schools, 48*(3), 168–182.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Hoboken, NJ: Wiley.
- Jenkins, J. R., & Johnson, E. (n.d.). Universal screening for reading problems: Why and how should we do this. *RTI action network*. Retrieved from <http://www.rtinetwork.org/essential/assessment/screening/readingproblems>
- Joanisse, M. F., Manis, F. R., Keating, P., & Seidenberg, M. S. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology, 77*, 30–60.
- Johnson, C. J., Beitchman, J. H., & Brownlie, E. B. (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*, 51–65.
- Johnson, E. S., Pool, J. L., & Carter, D. R. (2011). Validity evidence for the Test of Silent Reading Efficiency and Comprehension (TOSREC). *Assessment for Effective Intervention, 37*, 50–57.
- Kelso, K., Fletcher, J., & Lee, P. (2007). Reading comprehension in children with specific language impairment: An examination of two subgroups. *International Journal of Language & Communication Disorders, 42*, 39–57.
- Laing, G. J., Law, J., Levin, A., & Logan, S. (2002). Evaluation of a structured test and a parent led method for screening for speech and language problems: Prospective population based study. *BMJ, 325*, 1152. <https://doi.org/10.1136/bmj.325.7373.1152>
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*, 211–224.
- Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.) Cambridge, MA: MIT Press.
- Lipka, O., Lesaux, N. K., & Siegel, L. S. (2006). Retrospective analyses of the reading development of grade 4 students with reading disabilities: Risk status and profiles over 5 years. *Journal of Learning Disabilities, 39*, 364–378.
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*, 1198–1202.
- Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*, 1–14.
- Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). *Test of Silent Word Reading Fluency*. Austin, TX: Pro-Ed.
- McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the “specifics” of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry, 41*, 869–874.
- Mugnaini, D., Lassi, S., La Malfa, G., & Albertini, G. (2009). Internalizing correlates of dyslexia. *World Journal of Pediatrics, 5*, 255–264.
- National Center for Education Statistics. (2015). *Common core of data local education agency universe survey: School year 2013–2014 provisional data version 1a*. U.S. Department of Education. Washington, DC: Author. Retrieved from <https://nces.ed.gov/ccd/pubagency.asp>
- Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between

- poor reading comprehension and specific language impairment. *Journal of Speech, Language, and Hearing Research*, 47, 199–211.
- No Child Left Behind Act of 2001**, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).
- Plank, S. B., & Condliffe, B. F.** (2013). Pressures of the season: An examination of classroom quality and high-stakes accountability. *American Educational Research Journal*, 50, 1152–1182.
- Prelock, P. A., Hutchins, T., & Glascoe, F. P.** (2008). Speech-language impairment: How to identify the most common and least diagnosed disability of childhood. *The Medscape Journal of Medicine*, 10(6), 136.
- Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K.** (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain*, 136, 630–645.
- Ransby, M. J., & Swanson, H.** (2003). Reading comprehension skills of young adults with childhood diagnoses of dyslexia. *Journal of Learning Disabilities*, 36, 538–555.
- Reschly, D. J.** (2014). Response to intervention and the identification of specific learning disabilities. *Topics in Language Disorders*, 34, 39–58.
- Rice, M. L.** (2014). Advocating for SLI. Commentary on Bishop, D. V. M., 2014, Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders*, 49, 381–415.
- Rice, M. L.** (2016). Specific language impairment, nonverbal IQ, attention-deficit/hyperactivity disorder, autism spectrum disorder, cochlear implants, bilingualism, and dialectal variants: Defining the boundaries, clarifying clinical conditions, and sorting out causes. *Journal of Speech, Language, and Hearing Research*, 59, 122–132.
- SAS Institute Inc.** (2013). SAS/STAT® 9.4 Software. Cary, NC: SAS Institute Inc.
- Schatschneider, C.** (2013). I am ROC curves (and so can you). In Y. Petscher, C. Schatschneider, & D. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 65–92). New York, NY: Routledge.
- Semel, E., Wiig, E. H., & Secord, W. A.** (2004). *Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4)*. San Antonio, TX: Harcourt.
- Seymour, H. N., Roeper, T., & de Villiers, J. G.** (2003). *Diagnostic Evaluation of Language Variation—Screening Test*. Bloomington, MN: Pearson.
- Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D.** (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut Longitudinal Study. *Journal of the American Medical Association*, 264, 998–1002.
- Siegel, L. S.** (2008). Morphological awareness skills of English language learners and children with dyslexia. *Topics in Language Disorders*, 28, 15–27.
- Smolkowski, K., & Cummings, K. D.** (2015). Evaluation of the DIBELS diagnostic system for the selection of native and proficient English speakers at risk of reading difficulties. *Journal of Psychoeducational Assessment*, 34, 103–118.
- Snowling, M., Bishop, D. V. M., & Stothard, S. E.** (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry*, 41, 587–600.
- Snowling, M. J., Muter, V., & Carroll, J.** (2007). Children at family risk of dyslexia: A follow-up in early adolescence. *Journal of Child Psychology and Psychiatry*, 48, 609–618.
- South Carolina Department of Education.** (2017). *E-rate free and reduced meal eligibility data file 2013–2014*. Columbia, SC: South Carolina Department of Education. Retrieved from <http://ed.sc.gov/districts-schools/nutrition/meal-programs/national-school-lunch-program/e-rate-free-and-reduced-meal-eligibility-data/>
- Spencer, M., Quinn, J. M., & Wagner, R. K.** (2014). Specific reading comprehension disability: Major problem, myth, or misnomer? *Learning Disabilities Research & Practice*, 29, 3–9.
- Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M.** (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research*, 40, 1245–1260.
- Tomblin, J. B., Zhang, X., Buckwalter, P., & O'Brien, M.** (2003). The stability of primary language disorder four years after kindergarten diagnosis. *Journal of Speech, Language, and Hearing Research*, 46, 1283–1296.
- Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M.** (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45, 2–40.
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A.** (2010). *Test of Silent Reading Efficiency and Comprehension*. Austin, TX: Pro-Ed.
- Whitworth, A., Davies, C., Stokes, S., & Blain, T.** (1993). Identification of communication impairments in preschoolers: A comparison of parent and teacher success. *Australian Journal of Human Communication Disorders*, 21, 112–123.
- Wiig, E. H., Semel, E., & Secord, W. A.** (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)*. Bloomington, MN: NCS Pearson.
- Williams, K. T.** (2001). *Group Reading Assessment and Diagnostic Evaluation*. Circle Pines, MN: American Guidance Service.
- Wood, F. B.** (2013). *Predictive Assessment of Reading: Technical manual*. Winston Salem, NC: Wake Forest School of Medicine. Retrieved from <http://www.redesetgrow.com/wp-content/uploads/PAR2BTechManual.pdf>
- Wood, F. B., Hill, D. F., Meyer, M. S., & Flowers, D. L.** (2005). Predictive Assessment of Reading. *Annals of Dyslexia*, 55, 193–216.
- Woodcock, R. W.** (2011). *Woodcock Reading Mastery Tests—Third Edition (WRMT-III)*. Bloomington, MN: Pearson.

Appendix

Process of determining language impairment (LI) status for participants who spoke nonmainstream American English dialects (NMAE) versus mainstream American English dialects (MAE). Children with CELF-4 standard scores of ≤ 85 completed the DELV-ST to determine language variation status and obtain dialect-neutral screen of LI risk status, if needed. Children who spoke MAE according to the DELV-ST were classified as LI if they scored ≤ 1 SD below the mean on the CELF-4. Children who showed “some” or “strong” variation from MAE (DELV-ST Part I) were only classified as LI if (a) they also showed medium-high to high risk for LI (DELV-ST Part II) or (b) their CELF-4 standard score remained ≤ 85 after applying scoring modifications, as recommended by the CELF-4 manual. Scoring modifications were also applied for five individuals who did not complete the DELV-ST. CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition; DELV-ST = Diagnostic Evaluation of Language Variation–Screening Test; SS = standard score.