



Evaluation of the “Three Steps in Screening for Dyslexia” Assessment Protocol Designed for New Zealand Teachers

Mike Sleeman^{1,2} · John Everatt² · Alison Arrow² · Amanda Denston²

Received: 31 January 2022 / Accepted: 11 May 2022
© The Author(s) 2022

Abstract

Traditionally, the New Zealand Ministry of Education opposed the recognition of dyslexia. However, since 2007, the Ministry of Education’s position has started to change, evidenced by the development of a working definition. In 2021 the Ministry of Education released *Three Steps in Screening for Dyslexia* (TSSD), an assessment protocol designed to support teachers to screen for dyslexia. The current research evaluated the TSSD with a sample of 209 children in Years 4 to 6 (8–10 years-of-age) from New Zealand. The research investigated whether children could be accurately classified using tests from the TSSD, whether the three-step protocol described in the TSSD was a valid assessment approach, and what effect operationalising the term *average* at different cut-off points had on dyslexia screening. Children were classified using two cluster analyses. The first analysis was based on tests from the Woodcock Johnson IV and the second analysis was based on tests from the TSSD. Subsequent analyses investigated specific aspects of the TSSD protocol, including its sequential design and the placement of cut-off points. Results revealed a number of limitations to the TSSD approach. The authors discuss three changes that could be made to improve the validity and reliability of the TSSD, including a broader assessment of the decoding and language comprehension constructs; directing teachers to assess both decoding and language comprehension, irrespective of a child’s language comprehension ability; and placing a greater emphasis on discrepancy bands over cut-off points.

Keywords Dyslexia · Assessment · Specific comprehension difficulty · Simple view of reading · TSSD · Screening · Three Steps in Screening for Dyslexia

✉ Mike Sleeman
m.sleeman@massey.ac.nz

¹ College of Humanities and Social Sciences, Massey University, Albany Highway, Albany 0632, New Zealand

² College of Education, Health and Human Development, University of Canterbury, Christchurch, New Zealand

Introduction

New Zealand has a relatively large proportion of children who exhibit reading difficulties (Ministry of Education, 2017; Tunmer et al., 2013). The 2016 PIRLS data showed that New Zealand's mean reading scale score was worse than similar English-speaking countries from the OECD including Northern Ireland, the United States, Ireland, England, Canada, and Australia. Compared to these countries, a greater proportion of New Zealand's children also fell within the bottom two achievement bands (Ministry of Education, 2017). Some children within the lower end of the reading comprehension continuum exhibit difficulties consistent with a specific learning difficulty in reading, commonly called dyslexia. Traditionally, New Zealand has opposed the use of labels such as specific learning difficulty and dyslexia because of concerns that the use of labels may stigmatise some ethnicities who are more likely to exhibit reading difficulties (Tunmer & Chapman, 2007). However, groups advocating for those with dyslexia began to place increased pressure on the Ministry of Education to formally recognise dyslexia in the early 1990s (SPELD NZ, 2021).

The Ministry of Education formally recognised dyslexia in 2007 and a working definition was provided. This definition was controversial because it failed to include key components inherent within existing/accepted definitions of dyslexia. It also implied that children could have problems with musical notation, but no reading difficulties, yet still be considered as having dyslexia. This view is not consistent with accepted views within the scientific community (American Psychiatric Association, 2013; International Dyslexia Association, 2002; Rose, 2009). Tunmer and Greaney (2010) recommended a revised definition should be developed that defined dyslexia in terms of four key components: (a) persistent literacy difficulties (b) in otherwise typically developing children (c) despite exposure to high quality, evidence-based literacy instruction and intervention, (d) due to an impairment in the phonological processing skills required to learn to read and write. The Ministry of Education has since revised its definition to include these key components (Ministry of Education, 2016).

Whilst the Ministry of Education's definition of dyslexia is now consistent with contemporary research on dyslexia, considerable confusion remains regarding how classroom teachers can identify children with dyslexia. Nicholson and Dymock (2015) surveyed teachers on their ability to identify and support children with dyslexia. Most respondents (95%) believed they had children with dyslexia at their school, however, fewer believed they could identify these children (65%) and very few believed they were equipped to support children with dyslexia (12%). In recent years the Ministry of Education has come under considerable pressure to provide more specific guidance on how children with dyslexia can be identified and supported by classroom teachers (Ministry of Education, 2018).

Historically, New Zealand teachers have not been able to access valid and reliable tools that can be used to screen for dyslexia, meaning that assessment for dyslexia is typically conducted by a psychologist or a trained assessor associated with Specific Learning Difficulties New Zealand (New Zealand Qualifications

Authority, n.d.-a). Because these assessments are undertaken outside the school system, families are usually required to cover the assessment costs. Financial constraints have resulted in inequitable access to educational assessments (New Zealand Qualifications Authority, 2020). Assessors typically use the Wechsler scales or the Woodcock-Johnson batteries to test for dyslexia (New Zealand Qualifications Authority, n.d.-b) and most teachers are unable to administer tests from these batteries because they have not completed the necessary training to obtain the registration level required to purchase and administer these tests (New Zealand Qualifications Authority, n.d.-c).

In 2017, the government responded to 46 recommendations made by the Education and Science Select Committee report, which inquired into the identification of and support for children with dyslexia, dyspraxia, and autism spectrum disorders in primary and secondary schools (New Zealand Parliament, 2017). The recommendations included the need for earlier school-based identification of dyslexia and in 2019 the Ministry of Education released the *Learning Support Action Plan* (Ministry of Education, 2019) that identified six priorities designed to drive progress towards an inclusive education system. The second priority focused on screening and early intervention and stated that between July 2019 and December 2025 the Ministry of Education would develop evidence-based screening tools for dyslexia (Ministry of Education, 2019). In 2021, the Ministry of Education released *Three Steps in Screening for Dyslexia* (TSSD; Ministry of Education, 2021), a document that describes an assessment protocol that can be used by teachers to screen for dyslexia.

The TSSD is based on the Simple View of Reading (SVR; Gough & Tunmer, 1986), which predicts that children can be assigned to one of three poor reader groups based on their proficiency in decoding and language comprehension. Within this model, the label dyslexia is applied to children who exhibit decoding difficulties in the absence of language comprehension difficulties. The label specific comprehension difficulty (SCD) is applied to children who exhibit language comprehension difficulties in the absence of decoding difficulties and the label mixed difficulty is applied to children who exhibit both decoding and language comprehension difficulties. These three groups have been identified in previous SVR classification research (Aaron et al., 1999; Catts et al., 2003; Ebert & Scott, 2016; Morris et al., 2017) including a recent New Zealand study (Sleeman et al., 2021).

The TSSD directs teachers to follow a three-step protocol that begins with the assessment of a child's reading comprehension ability. If the child exhibits reading comprehension difficulties, teachers proceed to the second step of assessing language comprehension ability. According to the TSSD, children who exhibit average or above-average language comprehension ability are probably dyslexic and teachers are directed to the final step of assessing the child's decoding ability. The TSSD may not identify children who exhibit the mixed difficulty or SCD profiles because these children are likely to exhibit below-average language comprehension ability, which means their decoding ability will not be assessed. As a result, teachers will not obtain valuable information about children who exhibit reading comprehension difficulties that are not primarily due to decoding difficulties. This is particularly

troubling because recent research has identified that these groups may make up over 60% of poor readers (Sleeman et al., 2021).

The proportion of children who are identified as dyslexic using the TSSD may be influenced by the way teachers operationalise the term average. Reading comprehension, language comprehension, and decoding ability all fall along a continuum of performance, making it difficult to determine where cut-off points should be placed to discriminate between average and below-average performance. The placement of cut-off points on the reading comprehension and language comprehension variables will influence the proportion of children who progress through Steps 1 and 2 of the TSSD, which will influence the proportion of children who are identified as dyslexic. Teachers may interpret the term average in different ways, which will influence the proportion of children who are identified as dyslexic. For example, average could be interpreted as the median, so those below a 50% cut-off are 'below average'. It seems unlikely that many teachers would consider 50% of all children to have problems with reading, therefore, a lower cut-off is more useful. Those with more experience of assessment measures may have read about the 'average range', which is often taken as one standard deviation above and below the mean. This would lead to approximately the bottom 15% of scores being considered 'below average'. Other cut-off values have been used across a range of assessment measures, such as the bottom 10%, or the bottom 5%, and experience with these assessment tools may lead teachers to interpret below average in other ways. (See Sleeman, 2021, for a discussion of different cut-off values used in SVR validation studies.) Variation in the way teachers might operationalise the term average raises questions about the reliability of the TSSD for screening purposes.

The TSSD identifies a small number of informal and standardised tests that can be used by teachers and schools to assess reading comprehension, language comprehension, and decoding ability. The way reading comprehension, language comprehension, and decoding are defined and operationalised within the TSSD may influence the validity of the TSSD protocol for screening purposes. For example, within the SVR, reading comprehension is defined as the ability to extract meaning from linguistic discourse represented in print (Hoover & Tunmer, 2018). The key difference between reading comprehension and language comprehension is the medium in which these skills are applied. Reading comprehension focuses on text, whereas language comprehension focuses on speech. According to Hoover and Tunmer (2021) it is fundamental that parallel tests are used to assess these constructs, such as a passage read by an assessor followed by comprehension questions paired with a similar passage read by the child themselves followed by similar comprehension questions asked by the assessor. Varying only the medium of presentation enables conclusions to be made about whether the child's difficulties are primarily due to text reading difficulties or difficulties with understanding. However, the TSSD does not make clear the potential usefulness of matching the reading and listening comprehension measures used, which seems a further limitation of the current screening procedures. Given that measuring both modes of comprehension is indicated in the procedures, it seems a shame that the benefits of making these as equivalent as possible was not also detailed. This would not increase assessment procedures, but it may help teachers understand the logic behind such testing.

Under standardised assessment measures, the TSSD lists the PAT Reading Comprehension and PAT Listening Comprehension tests. However, these tests are not suitable for classification purposes because the tests report scale scores rather than standard scores. A relatively small change in scale scores on the Listening Comprehension test (around 2 scale scores) is equivalent to a year's worth of growth (New Zealand Council for Educational Research, 2014). This means it is difficult to discriminate between average and below-average performance within a year level. Raw scores on this test are also converted to stanines, which can be used to compare a child's performance to a national reference sample. Stanines span a range of percentiles so provide a more general indication of a child's performance relative to their peers than standard scores. This means they are likely to be less suitable for classification purposes than standard scores which map to an exact percentile. No other listening comprehension test is listed under standardised assessments in the TSSD; however, two expressive vocabulary tests are included (BPVS-III and the Peabody Picture Vocabulary Test). Assessing language comprehension ability using only a receptive vocabulary test may not provide a sufficiently broad assessment of the language comprehension construct. Assessment protocols that test children's vocabulary knowledge and listening comprehension ability provide a better indication of children's language comprehension ability than protocols that assess only one of these skills (Braze et al., 2007; Silverman et al., 2013). Nevertheless, the TSSD does not direct teachers to assess both of these skills.

Care must also be taken when operationalising the decoding construct. Decoding, within the SVR, is defined as the ability to quickly, accurately, and effortlessly access word meanings from our mental lexicon (Hoover & Tunmer, 2018). It is typically operationalised using word identification and word-attack tests (Adlof et al., 2006; Catts et al., 2006; Language & Reading Research Consortium, 2015). The former typically involves reading aloud individual real words of varying complexity (such as varying frequency of familiarity or age of acquisition) out of context (i.e., not in a sentence) whereas the latter typically involves pronouncing made-up words or non-words/pseudo-words, which provides an assessment of how the child deals with novel letter strings. The TSSD identifies both word identification (Burt Word Recognition Test and STAR Test) and word attack (Martin & Pratt Non-Word Reading Test) tests, however, the word attack test listed (Martin & Pratt Non-Word Reading Test) is no longer sold by the New Zealand Council for Educational Research (NZCER), the organisation that sells educational assessments to New Zealand schools, meaning that teachers may forgo assessing this skill. Yet, assessing only word identification ability provides too narrow an assessment of the decoding construct. The ability to decode words using knowledge of phoneme-grapheme relationships and English spelling rules is a prerequisite for orthographic mapping (Frith, 1980; Tunmer & Hoover, 2019), thus it is important for teachers to screen a child's word attack ability because it is a primary feature associated with difficulties in learning to read and spell presented by those with dyslexia (Everatt & Denston, 2020; Gillon, 2018; Nicholson & Dymock, 2015; Snowling, 2000).

This review of the TSSD identified three limitations that may influence the validity and reliability of the assessment protocol: (a) research has not confirmed whether a sequential three-step assessment protocol is an appropriate screening method for

dyslexia; (b) teachers are asked to differentiate between average and below-average performance but the TSSD does not describe how the term average should be operationalised; (c) research has not confirmed whether the tests described in the TSSD provide a valid assessment of the reading comprehension, language comprehension, and decoding constructs.

The aforementioned limitations mean that research is needed to investigate whether the TSSD is a valid dyslexia screening protocol. The present research examined whether classification based on tests from the TSSD can identify children who exhibit a dyslexia profile. It also investigated what impact the use of a three-step protocol has on the identification and classification of children with reading difficulties. Finally, it examined what proportion of children is identified as dyslexic when using different language comprehension cut-off points. Having determined a liberal cut-off for reading comprehension weaknesses to avoid missing children with reading difficulties, the classification of difficulties into dyslexia-based versus language comprehension still requires a cut-off for the second step, which assesses children's language comprehension ability. In this study, we look at the impact using various points for this second step decision has, given that the majority of those with reading comprehension weaknesses should be in the sample.

Method

Participants

The participants came from nine primary schools in an urban city in New Zealand. These children were in Years 3, 4, and 5 (aged 8–10 years). Children in these year levels were targeted as reading comprehension ability is influenced, to a similar extent, by both decoding and language comprehension ability in this age range (Adlof et al., 2006; Catts, 2018; Georgiou et al., 2009). Principals and teachers of the target year groups were provided with information and consent forms and all agreed to support the research by identifying children to participate in the research and releasing them from their classrooms to complete the four individually administered assessment sessions. Schools were asked to identify children who performed below the 40th percentile on one of two school-based standardised assessments commonly used within New Zealand: the e-asTTle Reading test (Auckland UniServices Limited, 2009) or the Progressive Achievement Test for Reading Comprehension (Darr et al., 2008). A liberal cut-off point (40th percentile) was used in this research because New Zealand has a relatively large proportion of children who exhibit reading difficulties (Ministry of Education, 2017). The 40% figure was considered to be the most liberal that a teacher would use (see above) and would include those who would have been selected via a lower percentile score. Teachers were also able to nominate children who exhibited reading difficulties on other school assessments. All of the children identified were invited to take part in this research. Consent to participate in this research was obtained from all the participating children and their parents. This research adhered to the ethical requirements of the New

Zealand university in which the authors were working and an application to one of its ethics committees was approved.

In total, 216 English-speaking children took part in this study. Seven children performed above the 40th percentile on the researcher administered Passage Comprehension test from the Woodcock-Johnson IV (WJIV; Schrank et al., 2014) and were excluded from the research, leaving a final sample of 209 children with an average age of nine years and eight months ($SD_{\text{age}} = 11$ months). Table 1 provides an overview of the participants broken down by year and gender.

Procedure and Measures

All children undertook seven individually administered assessments carried out by the first author. The assessments, for each child, were completed over four sessions lasting approximately 20 min each within a two-week period. For reliability purposes, a second marker reviewed 20% of the assessment record sheets. No discrepancies between markers were identified during this process. Decoding and language comprehension were assessed using tests from the WJIV that are commonly used by assessors (New Zealand Qualifications Authority, n.d.-b) and researchers (Aaron et al., 1999; Catts et al., 2003) who wish to identify children with dyslexia. Decoding and language comprehension were also assessed using tests named for use in the TSSD. The following sections describe these tests. Tests that assess decoding and language comprehension ability from the WJIV that are typically used by assessors were included to determine the conclusions from a professional assessment. Tests from the recommended list in the TSSD were used to indicate what a teacher using this protocol may conclude. Comparisons of the two with the same children can then be used to determine the level of similarity/dissimilarity between the two procedures.

WJIV Reading Comprehension Assessment

Reading comprehension ability was assessed using the Passage Comprehension test from the WJIV (Schrank et al., 2014). This test required students to read short passages of text silently and then supply a key missing word in each passage. The initial items on this test were one sentence in length. As children progressed through the test, the items increased in length and complexity. The *Examiners Manual* (Mather

Table 1 Participant demographics

Year	Males <i>n</i> (% of gender)	Females <i>n</i> (% of gender)	Total <i>n</i> (% of all participants)
4	35 (62.5%)	21 (37.5%)	56 (26.8%)
5	49 (68.1%)	23 (31.9%)	72 (34.4%)
6	46 (56.8%)	35 (43.2%)	81 (38.8%)
Total	130 (62.2%)	79 (37.8%)	209 (100.0%)

& Wendling, 2014) reports median reliability of .89 for the Passage Comprehension test within the 5–19 age range.

WJIV Decoding/Word Reading Assessments

Decoding ability was assessed using the Letter-Word Identification and Word Attack tests from the WJIV (Schrank et al., 2014). The Letter-Word Identification test assessed children's ability to identify and pronounce individual letters and words. The Word Attack test assessed children's ability to pronounce non-words that conform to English spelling rules. The Letter-Word Identification ($r = .95$) and Word Attack ($r = .87$) tests demonstrated excellent reliability within this sample, similar to those reported in the WJIV manual (Schrank et al., 2014; Letter-Word Identification = .92, Word Attack = .90). Both tests were administered following the procedures described in the WJIV manual and were stopped when a child made six consecutive errors.

TSSD Decoding/Word Reading Assessment

The Burt Word Recognition Test (Burt test; Gilmore et al., 1981) from the TSSD was used to assess children's word identification ability. This test assessed children's ability to read a range of regular and irregular words that increased in length and complexity. Testing ceased when children were unable to correctly read 10 consecutive items. The test manual reports high internal consistency (.97) within the 8.03–10.09 age range (Gilmore et al., 1981). Word-attack was not assessed as the TSSD test identified to assess word attack ability is not available for teachers.

WJIV Language Comprehension

Language comprehension was measured using the Oral Comprehension test from the WJIV Oral Language battery and the Oral Vocabulary test from the WJIV Cognitive battery. The Oral Comprehension test required children to listen to short passages and then supply a missing final word to each. The Oral Vocabulary test required children to provide synonyms and antonyms for orally presented words. The tests demonstrated excellent reliability within this sample (Oral Comprehension = .75; Oral Vocabulary = .84), similar to those reported in the WJIV manual (Schrank et al., 2014; Oral Comprehension = .82, Oral Vocabulary = .89). The tests were administered following the procedures outlined in the WJIV manual and ceased when the child made six consecutive errors.

TSSD Language Comprehension Assessment

The TSSD identifies two standardised tests that can be used to assess vocabulary ability (British Picture Vocabulary Scale and Peabody Picture Vocabulary Test). In the current research, children's vocabulary ability was assessed using the British Picture Vocabulary Scale, 3rd Edition (BPVS-III; Dunn et al., 2009). The BPVS-III

test assessed children's ability to identify one picture from a selection of four that represented an orally presented word. A reliability figure of 0.91 has been reported for this assessment (Dockrell & Marshall, 2015). This test was discontinued once children made eight or more errors in a set of 12 items. The listening comprehension test (PAT Listening Comprehension) was not administered because it is difficult to differentiate accurately between average and below-average performance within a year level using the scale scores and stanines reported by this assessment.

Standard and Composite Scoring

Initially, raw scores from the WJIV subtests and the BPVS-III were converted to standard scores using the relevant administration manuals or conversion software. The Burt test reports age equivalent bands. Whilst these scores can be used to identify average and below-average performance, they were not suitable for classification purposes in this research. We calculated standard scores for the Burt test using the mean and standard deviation from a study that administered this assessment to a similar age group of children who were representative of all ability levels (see Mandelaine & Wheldall, 1998). A composite decoding score was calculated by finding each child's average standard score on the Word Attack and Letter Word Identification tests. A composite language comprehension score was derived by calculating each child's average score on the Oral Comprehension and Oral Vocabulary tests. These composite scores were used to classify children. To ensure the four aforementioned tests provided a reliable indication of children's decoding and language comprehension ability, reliability scores were calculated for this sample and have been reported within the above sections. All analyses were performed using IBM SPSS Statistics for Windows, version 26.0.

Results

Table 2 reports the mean and standard deviation in standard scores for each test that was administered.

Table 2 Descriptive statistics

Test	Construct	Total <i>M</i> (<i>SD</i>)
Passage comprehension ^a	Reading comprehension	77.87 (11.42)
Word attack ^a	Decoding	84.90 (14.68)
Letter-word identification ^a	Decoding	84.14 (13.85)
Burt word reading ^b	Decoding	85.15 (11.28)
Oral comprehension ^a	Language comprehension	85.84 (12.97)
Oral vocabulary ^a	Language comprehension	82.13 (13.40)
BPVS-III ^b	Language comprehension	81.54 (9.70)

^aTest from the WJIV

^bTest from the TSSD

Initially, children were classified according to their performance on the composite decoding and language comprehension variables. This analysis is referred to as the WJIV classification analysis. The composite variables were entered into a two-step cluster analysis that used log-likelihood as the distance measure. In the first step, the program examined every record and decided whether that record should be merged with a previously formed group of records (cluster) or whether it should form the basis for a new cluster based on a specified distance criterion. In the second step, the program took the clusters identified in the first step and grouped them into the desired number of clusters. In this analysis, the program was allowed to determine the optimal number of groupings. Recent research indicated that classification based on a cluster analysis provided a better fit for the data than alternative cut-off line approaches (Sleeman et al., 2021). As expected, three poor reader groups were identified: mixed difficulty, SCD, and dyslexia. The majority of children (83.3%) were assigned to the dyslexia (38.8%) and SCD (44.5%) groups. A smaller proportion of children were assigned to the mixed difficulty group (16.7%). Table 3 reports the proportion of children who were assigned to each group.

The next analyses investigated whether the Burt and BPVS-III tests from the TSSD can be used for classification purposes. Analyses indicated a strong correlation between the Burt test and the decoding variable ($r = .767$, $N = 209$, $p < .001$) and between the BPVS-III and the language comprehension variable ($r = .624$, $N = 209$, $p < .001$). This confirms the predicted positive relationship between these variables, however, it does not confirm that these tests will lead to the same classification outcomes.

To evaluate whether the same three poor reader groups can be identified using the Burt and BPVS-III tests a second classification analysis was conducted. As in the previous classification analysis, children were classified using a two-step cluster analysis, referred to as the TSSD classification analysis. Log likelihood was again used as the distance measure and the analysis software program was allowed to determine the optimal number of groupings. The classification analysis using the TSSD identified the same poor reader groups predicted by the SVR. However, the proportion of children assigned to each group differed across the WJIV classification analysis and TSSD classification analysis. A larger proportion of children were assigned to the mixed difficulty group in the TSSD classification analysis (35.9%) than the WJIV classification analysis (16.7%) and a smaller proportion of children were assigned to the dyslexia group (19.6% and 38.8% respectively). The same proportion of children was assigned to the SCD group across the classification approaches (44.5%). Table 3 reports the proportion of children assigned to each group across the classification approaches.

Table 3 Classification analysis

Approach	Dyslexia % (<i>n</i>)	SCD % (<i>n</i>)	Mixed difficulty % (<i>n</i>)
WJIV classification analysis	38.8% (81)	44.5% (93)	16.7% (35)
TSSD classification analysis	19.6% (41)	44.5% (93)	35.9% (75)

Table 4 Assignment comparison between the WJIV classification analysis and the TSSD classification analysis

	TSSD classification analysis		
	Mixed (%)	SCD	Dyslexia
WJIV classification analysis			
Mixed	89	11	0
SCD	19	71	10
Dyslexia	32	28	40

Table 4 reports the results from an analysis that examined the individual-level assignment to groups across the WJIV classification analysis and the TSSD classification analysis. If children were assigned to the same groups across the analyses it would indicate that, respectively, the Burt and BPVS-III tests provide an accurate assessment of decoding and language comprehension ability, which would mean they may be suitable for classification purposes. The results indicate that most children in the mixed (89%) and SCD (71%) groups were assigned to the same group across these approaches. In contrast, a relatively small proportion of children (40%) were accurately assigned to the dyslexia group in the TSSD classification analysis.

The TSSD recommends assessing decoding ability only if a child performs within the average or above-average range on the language comprehension variable. This analysis investigated what effect operationalising average language comprehension ability at different cut-off points (10th, 20th, 30th, 40th, and 50th percentile) had on the proportion of children who were identified as dyslexic using data from the WJIV classification analysis. Note that 81 children were identified as dyslexic in the cluster analysis using scores on the WJIV. However, one of these 81 children performed below the 10th percentile on the language comprehension test. If we use the 10th percentile as the cut-off point in the screening procedures, this child would be considered non-dyslexic and would be excluded from the procedures at this point. However, it seems unlikely that a teacher would interpret the 10th percentile as the point to determine average or above-average skills. To examine this, analyses were performed to compare different cut-offs. These analyses used data from the WJIV classification analysis, rather than the TSSD analysis because of the limitations associated with the TSSD classification analysis. When stricter cut-off points were used (> 10th percentile), fewer children were identified as dyslexic because they did not

Table 5 Proportion of children identified as dyslexic using different language comprehension cut-off points

Language comprehension cut-off point	Proportion of dyslexic students	
	Correctly identified	Not identified
50th	16 (20%)	65 (80%)
40th	24 (30%)	57 (70%)
30th	37 (46%)	44 (54%)
20th	68 (84%)	13 (16%)
10th	80 (99%)	1 (1%)
No cut-off point	81 (100%)	0 (0%)

perform above the language comprehension cut-off point. The results indicated (see Table 5) that operationalising average performance as performance above the 40th percentile led to 70% of children who exhibited the dyslexic profile in the WJIV cluster analysis now being classified as non-dyslexic. This is despite these children showing difficulties with reading comprehension and decoding in the WJIV data. And note that these children's decoding difficulties would not be assessed if this cut-off were used in the TSSD procedures. Such children would be classified as not needing dyslexia-related support only because they are showing some weaknesses in the listening comprehension measure, weaknesses that may be due to poor reading experience leading to less exposure to words, which can in turn lead to poorer vocabulary levels that can impact on listening comprehension performance.

Discussion

Although the TSSD classification analysis also identified the three poor reader groups identified by the WJIV assessment measures, the results indicated that children who exhibit a dyslexia profile were not consistently identified across the two analyses. This suggests that teacher-implemented TSSD procedures will produce differing dyslexia-identification results to assessment procedures that use more widely used standardised tests, and that the recommended, and available, word identification and receptive vocabulary tests may provide an incomplete assessment of the decoding and language comprehension constructs.

The findings indicated that only 40% of children identified as having dyslexia by the WJIV assessment measures were assigned to the dyslexia group in the TSSD classification analysis. Over 30% of the children identified as having dyslexia via the WJIV assessment measures were assigned to the mixed difficulty group using the TSSD method. This indicates that when they were assessed using a receptive vocabulary test many of the children showed poorer scores on the language comprehension measure in the TSSD procedure. Difficulties with vocabulary may be due to a lack of reading experience leading to poorer exposure to words in differing contexts that receptive vocabulary measures assess. This negative impact may then increase with age as those with poor reading experience encounter fewer word-context exposures compared to those with good reading levels (i.e., Matthew effects as suggested by Stanovich, 1986).

Nearly 30% of children who were assigned to the dyslexia group by the WJIV assessment measures were assigned to the SCD group in the TSSD classification analysis. This movement is surprising because the dyslexia and SCD groups should exhibit opposing profiles. Children with dyslexia should show decoding difficulties along with relatively unaffected language comprehension ability; whereas, children who exhibit the SCD profile should demonstrate language comprehension difficulties with relatively unaffected decoding ability. To move from the dyslexia group in the WJIV classification analysis to the SCD group in the TSSD classification analysis, children must have performed better on the Burt test than the composite decoding variable and worse on the BPVS-III than the composite language comprehension variable. The latter suggests that these

children find receptive vocabulary tests more difficult than tests that assess listening comprehension and expressive vocabulary ability as discussed in the previous paragraph. However, this finding suggests that those with dyslexia will find word identification tests relatively easy compared to word attack tests. This may be due to some children being able to recall words that they have been taught easier than having to use decoding processes to translate letter strings into appropriate verbal forms.

Reading Comprehension

The current research asked schools to identify children who were exhibiting reading difficulties using the e-asTTle Reading test, the Progressive Achievement Test for Reading Comprehension, or other classroom reading assessments. Nearly 97% of children who were nominated by schools performed below the 40th percentile on the researcher administered reading comprehension test. This suggests that teachers could use classroom assessment data to identify children with reading difficulties. Whilst this is a promising finding, we cannot rule out the possibility that other children in these classrooms should have been identified as struggling readers, as we did not assess children who were not nominated to participate in this research.

Language Comprehension

The TSSD states that after assessing reading comprehension ability teachers should assess children's language comprehension ability (Ministry of Education, 2021). One of the recommended language comprehension tests, the BPVS-III, was used in this research. Results indicated that assessing only receptive vocabulary ability, as the BPVS-III does, did not provide a sufficiently robust assessment of language comprehension ability. The TSSD identifies two additional standardised tests that can be used to assess language comprehension ability: The Peabody Picture Vocabulary Test (PPVT) and the PAT Listening Comprehension test. Both the PPVT and the BPVS-III assess receptive vocabulary ability so it seems unlikely that the PPVT will provide a more robust assessment of language comprehension ability than the BPVS-III. Listening comprehension ability was not assessed because the PAT Listening Comprehension test is not sufficiently accurate for classification purposes (New Zealand Council for Educational Research, 2014). The results from the present analyses indicate that assessing listening comprehension ability, in addition to vocabulary knowledge, provides a more robust assessment of language comprehension than vocabulary alone. To ensure listening comprehension is assessed, an appropriate listening comprehension test should be added to the TSSD. Furthermore, the TSSD should also state that teachers should assess both vocabulary and listening comprehension ability. Assessing only one of these skills may not provide a sufficiently accurate assessment of a child's language comprehension ability.

Decoding

If a teacher finds that a child exhibits reading comprehension difficulties but has average or above-average language comprehension ability, the TSSD states that decoding ability should be assessed (Ministry of Education, 2021). The Ministry of Education identifies three standardised tests that can be used to assess decoding ability. One of these tests is the Burt test, which was used to assess word identification ability in this research. Whilst there was a strong positive correlation between the Burt and composite decoding variable, analyses indicated that this test did not provide a sufficiently broad assessment of decoding ability. In the present study, some children found the Word Attack test from the WJIV particularly difficult, which suggests that assessing decoding ability requires using both word identification and word attack tests. The TSSD does not identify a standardised word attack test that can be purchased through NZCER, for use in New Zealand schools. It is recommended that a standardised word attack test that can be purchased from NZCER should be added to the TSSD and the TSSD should state that teachers must assess both a child's word identification and word attack ability. Both word and non-word reading measures can be quick to administer, particularly those with a good stop rule (i.e., the point when the test is stopped due to the number of errors made), so including both in a screening procedure would not increase testing time substantially but should increase the accuracy of assessment and hence improve support.

Assessment Protocol

The TSSD (Ministry of Education, 2021) states that decoding ability should be assessed if a child exhibits reading comprehension difficulties and average, or above average, language comprehension ability. Because struggling readers can exhibit the dyslexia, mixed difficulty, or SCD profile, both decoding and language comprehension ability should be assessed when reading comprehension difficulties are identified. Children who exhibit decoding difficulties in addition to language comprehension difficulties exhibit the mixed difficulty profile. Whereas, children who exhibit unaffected decoding ability exhibit the SCD profile. These children will likely benefit from different instructional approaches because the root cause of their reading difficulties varies (Aaron et al., 2008; Carson et al., 2013; Clarke et al., 2010; Gillon et al., 2019).

The results indicate that the TSSD should define average ability. The proportion of children who were identified as dyslexic varied considerably depending on how the term was operationalised. Teachers may subjectively define this term if a clear definition is not provided, thus potentially reducing the consistency of the tool across schools and between teachers. Rather than directing teachers to screen students using cut-off points, the TSSD could direct teachers to look for a discrepancy between a child's decoding and language comprehension ability. This approach is similar to that employed in the WJIV classification analysis, which identified a group of dyslexic children who exhibited decoding difficulties and relatively unaffected language comprehension ability. If this approach is adopted, the TSSD should

describe how a discrepancy should be operationalised. It may be helpful to use discrepancy bands. For example, a large discrepancy may place a child within the high-risk range for dyslexia. However, smaller discrepancies may place a child within a moderate or low-risk group. Because the discrepancy between children's decoding and language comprehension ability falls along a continuum, the use of discrepancy bands may provide a more accurate screening approach than approaches that rely on a single cut-off point. Similar approaches using significant differences (Sleeman, 2021) and discrepancies based on standard deviations (Wagner et al., 2020) have been used to identify children who exhibit the dyslexia profile.

Limitations and Future Research Opportunities

The relative importance of decoding and language comprehension to reading comprehension changes over time (Catts, et al., 2005; Georgiou et al., 2009; Hoover & Gough, 1990), which influences the proportion of children who are assigned to the dyslexia, SCD, and mixed difficulty groups (Catts et al., 2003). Future research examining whether similar patterns to those found in the current research can be identified at different year levels would be valuable as part of comprehensive recommendations across New Zealand school years. Such research may also find that a different set of tests would be more useful to identify the dyslexia, SCD, and mixed difficulty profiles at higher year levels. For example, some research has found that fluency measures provide a better indication of decoding ability in older children than accuracy-only measures (Kershaw & Schatschneider, 2012).

Conclusion

The results indicate that the TSSD requires further refinement to be suitable for the screening of dyslexia and other reading difficulties. Although the process has some positive features, particularly in terms of increasing awareness of issues related to dyslexia and poor reading comprehension among mainstream New Zealand teachers, it has three main limitations. First, the decoding and language comprehension tests from the TSSD do not provide a sufficiently broad assessment of these constructs. As a result, some children (particularly those with dyslexia) may be misclassified using these tests. Second, once teachers identify a child who exhibits reading comprehension difficulties they should be directed to assess both the child's decoding and language comprehension ability. This information can be used to discriminate between children who exhibit the dyslexia, SCD, or mixed difficulty profiles, which may then inform initial support strategies (e.g., focus on making the link between writing and language sounds and/or improving vocabulary knowledge and strategies for understanding oral discourse). Finally, encouraging teachers to identify at-risk children using discrepancy bands based on children's decoding and language comprehension ability may lead to fewer false positives and false negatives than the current approach of determining average performance. With a strategically selected/developed set of tests, these bands could even be provided for teachers. Therefore,

although partially negative, the findings from this study should inform the development of a more accurate dyslexia screening tool suitable for use within New Zealand schools.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aaron, P. G., Joshi, M., & Williams, K. A. (1999). Not all reading disabilities are alike. *Journal of Learning Disabilities*, 32(2), 120–137. <https://doi.org/10.1177/002221949903200203>
- Aaron, P. G., Joshi, R. M., Gooden, R., & Bentum, K. E. (2008). Diagnosis and treatment of reading disabilities based on the component model of reading: An alternative to the discrepancy model of LD. *Journal of Learning Disabilities*, 41(1), 67–84. <https://doi.org/10.1177/0022219407310838>
- Adlof, S., Catts, H., & Little, T. (2006). Should the simple view of reading include a fluency component? *Reading and Writing*, 19(9), 933–958. <https://doi.org/10.1007/s11145-006-9024-z>
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Publishing.
- Auckland UniServices Limited. (2009). *e-asTTle reading*. Ministry of Education.
- Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities*, 40(3), 226–243. <https://doi.org/10.1177/00222194070400030401>
- Carson, K. L., Gillon, G. T., & Boustead, T. M. (2013). Classroom phonological awareness instruction and literacy outcomes in the first year of school. *Language, Speech, and Hearing Services in Schools*, 44(2), 147–160. [https://doi.org/10.1044/0161-1461\(2012/11-0061\)](https://doi.org/10.1044/0161-1461(2012/11-0061))
- Catts, H. (2018). The simple view of reading: Advancements and false impressions. *Remedial and Special Education*, 39(5), 317–323. <https://doi.org/10.1177/0741932518767563>
- Catts, H., Adlof, S., & Weismer, S. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, 49(2), 278–293. [https://doi.org/10.1044/1092-4388\(2006/023\)](https://doi.org/10.1044/1092-4388(2006/023))
- Catts, H., Hogan, T., & Adlof, S. (2005). Developmental changes in reading and reading disabilities. In H. Catts & A. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Lawrence Erlbaum Associates.
- Catts, H., Hogan, T., & Fey, M. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities*, 36(2), 151–164. <https://doi.org/10.1177/002221940303600208>
- Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, 21(8), 1106–1116. <https://doi.org/10.1177/0956797610375449>
- Darr, C., Ferral, H., Twist, J., & Watson, V. (2008). *PAT (Progressive Achievement Test) – Reading comprehension: Revised 2008*. NZCER.
- Dockrell, J. E., & Marshall, C. R. (2015). Measurement Issues: Assessing language skills in young children. *Child and Adolescent Mental Health*, 20(2), 116–125. <https://doi.org/10.1111/camh.12072>

- Dunn, L., Dunn, D., Sewell, J., Styles, B., Brzyska, B., Shamsan, Y., & Burge, B. (2009). *The British picture vocabulary scale third edition* (3rd ed.). GL Education.
- Ebert, K. D., & Scott, C. M. (2016). Bringing the simple view of reading to the clinic: Relationships between oral and written language skills in a clinical sample. *Journal of Communication Disorders*, 62, 147–160. <https://doi.org/10.1016/j.jcomdis.2016.07.002>
- Everatt, J., & Denston, A. (2020). *Dyslexia: Theory, assessment and support*. Routledge.
- Frith, A. (1980). *Cognitive processes in spelling*. Academic Press.
- Georgiou, G. K., Das, J. P., & Hayward, D. (2009). Revisiting the “simple view of reading” in a group of children with poor reading comprehension. *Journal of Learning Disabilities*, 42(1), 76–84. <https://doi.org/10.1177/0022219408326210>
- Gillon, G. T. (2018). *Phonological awareness: From research to practice* (2nd ed.). Guilford Press.
- Gillon, G., McNeill, B., Scott, A., Denston, A., Wilson, L., Carson, K., & Macfarlane, A. H. (2019). A better start to literacy learning: Findings from a teacher-implemented intervention in children’s first year at school. *Reading & Writing*, 32(8), 1989–2012. <https://doi.org/10.1007/s11145-018-9933-7>
- Gilmore, A., Croft, C., & Reid, N. (1981). *Burt Word Recognition Test: New Zealand revision*. New Zealand Council for Educational Research.
- Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10. <https://doi.org/10.1177/074193258600700104>
- Hoover, W., & Gough, P. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. <https://doi.org/10.1007/BF00401799>
- Hoover, W., & Tunmer, W. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education*, 39(5), 304–312. <https://doi.org/10.1177/0741932518773154>
- Hoover, W. A., & Tunmer, W. E. (2021). The primacy of science in communicating advances in the science of reading. *Reading Research Quarterly*. <https://doi.org/10.1002/rrq.446>
- International Dyslexia Association. (2002). *Definition consensus project*. <https://dyslexiaida.org/definition-consensus-project/>
- Language and Reading Research Consortium. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, 50(2), 151–169. <https://doi.org/10.1002/rrq>
- Mandelaine, A., & Wheldall, K. (1998). Towards a curriculum-based passage reading test for monitoring the performance of low-progress readers using standardised passages: A validity study. *Educational Psychology*, 18(4), 471–498. <https://doi.org/10.1080/0144341980180408>
- Mather, N., & Wendling, B. (2014). *Examiners manual: Woodcock Johnson IV Test of Cognitive Abilities*. Riverside.
- Ministry of Education. (2016). *Defining dyslexia*. http://literacyonline.tki.org.nz/Literacy-Online/Learning-about-my-students-needs/Knowledge-of-the-learner/Literacy-and-students-with-special-education-needs#eztoc87311_0_1
- Ministry of Education. (2017). *PIRLS 2016 New Zealand’s achievement*. <https://www.educationcounts.govt.nz/publications/schooling2/large-scale-international-assessments/pirls-201516>
- Ministry of Education. (2018). *Draft disability and learning support action plan. Analysis of engagement and feedback*. <https://conversation.education.govt.nz/assets/DLSAP/Draft-Disability-and-Learning-Support-Action-Plan-Analysis-of-Engagement-Feedback.pdf>
- Ministry of Education. (2019). *Learning Support Action Plan 2019–2025*. <https://conversation.education.govt.nz/assets/DLSAP/Learning-Support-Action-Plan-2019-to-2025-English-V2.pdf>
- Ministry of Education. (2021). *Three steps in screening for dyslexia*. <https://inclusive.tki.org.nz/guides/dyslexia-and-learning/classroom-assessments-and-monitoring/>
- Morris, D., Meyer, C., Trathen, W., McGee, J., Vines, N., Stewart, T., Gill, T., & Schlagal, R. (2017). The simple view, instructional level, and the plight of struggling fifth-/sixth-grade readers. *Reading & Writing Quarterly*, 33(3), 278–212. <https://doi.org/10.1080/10573569.2016.1203272>
- New Zealand Council for Educational Research. (2014). *PAT info sheet*. <https://www.nzcer.org.nz/system/files/PAT%20Info%20sheet%202014.pdf>
- New Zealand Parliament. (2017). *Government response to the Report of the Education and Science Select Committee, on its inquiry into the identification and support for students with the significant challenges of dyslexia, dyspraxia and autism spectrum disorders in primary and secondary schools: Presented to the House of Representatives in accordance with Standing Order 252*. https://www.parliament.nz/en/pb/papers-presented/current-papers/document/PAP_72401/government-response-to-report-of-the-education-and-science

- New Zealand Qualifications Authority. (n.d-a). *SAC information for schools*. <https://www.nzqa.govt.nz/providers-partners/assessment-and-moderation-of-standards/managing-national-assessment-in-schools/special-assessment-conditions/sac-information-for-schools/evidence-needed/>
- New Zealand Qualifications Authority. (n.d.-b). *Screening for entitlement to special assessment conditions*. <https://www.nzqa.govt.nz/audience-pages/secondary-schools-and-teachers/forms/>
- New Zealand Qualifications Authority. (n.d.-c). *Special assessment conditions - information for registered professionals*. <https://www.nzqa.govt.nz/providers-partners/assessment-and-moderation-of-standards/managing-national-assessment-in-schools/special-assessment-conditions/for-professionals/who-may-provide-evidence-in-support-of-a-students-sac-needs/>
- New Zealand Qualifications Authority. (2020). *NCEA special assessment conditions 2018, 2019* (CR20856). <https://www.nzqa.govt.nz/about-us/oia-requests/information-releases/>
- Nicholson, T., & Dymock, S. (2015). *The New Zealand dyslexia handbook*. NZCER Press.
- Rose, J. (2009). *Identifying and teaching children and young people with dyslexia and literacy difficulties: An independent report*. <https://dera.ioe.ac.uk/14790/>
- Schrank, F., McGrew, K., & Mather, N. (2014). *Woodcock Johnson IV*. Riverside.
- Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading, 17*(2), 108–133. <https://doi.org/10.1080/10888438.2011.618153>
- Sleeman, M. (2021). *The identification and classification of reading disorders based on the simple view of reading* [Doctoral thesis, University of Canterbury]. <https://go.exlibris.link/q5tzNsXQ>
- Sleeman, M., Everatt, J., Arrow, A., & Denston, A. (2022). *The identification and classification of struggling readers based on the simple view of reading*. Manuscript in preparation.
- Snowling, M. J. (2000). *Dyslexia* (2nd ed.). Blackwell.
- SPELD NZ. (2021). 50 years breaking barriers. *The Record, 8*, 1–4.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360–407. <https://doi.org/10.1598/RRQ.21.4.1>
- Tunmer, W., & Chapman, J. (2007). Language-related differences between discrepancy-defined and non-discrepancy-defined poor readers: A longitudinal study of dyslexia in New Zealand. *Dyslexia, 13*(1), 42–66. <https://doi.org/10.1080/10349120120115307>
- Tunmer, W., Chapman, J., Greaney, K., Prochnow, J., & Arrow, A. (2013). Why the New Zealand National Literacy Strategy has failed and what can be done about it: Evidence from the Progress in International Reading Literacy Study (PIRLS) 2011 and reading recovery monitoring reports. *Australian Journal of Learning Difficulties, 18*(2), 139–180. <https://doi.org/10.1080/19404158.2013.842134>
- Tunmer, W., & Greaney, K. (2010). Defining dyslexia. *Journal of Learning Disabilities, 43*(3), 229–243. <https://doi.org/10.1177/0022219409345009>
- Tunmer, W., & Hoover, W. (2019). The cognitive foundations of learning to read: A framework for preventing and remediating reading difficulties. *Australian Journal of Learning Difficulties, 24*(1), 75–93. <https://doi.org/10.1080/19404158.2019.1614081>
- Wagner, R. K., Zirps, F. A., Edwards, A. A., Wood, S. G., Joyner, R. E., Becker, B. J., Liu, G., & Beal, B. (2020). The prevalence of dyslexia: A new approach to its estimation. *Journal of Learning Disabilities, 53*(5), 354–365. <https://doi.org/10.1177/0022219420920377>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.