DYSLEXIA
Published online in Wiley Online Library
(wileyonlinelibrary.com). DOI: 10.1002/dys.1492

A New Self-Report Inventory of Dyslexia for Students: Criterion and Construct Validity

Peter Tamboer[1]* and Harrie C. M. Vorst[2]

[1]*University of Amsterdam, Department of Psychological Methods, Weesperplein 4, Room 218, 1018 XA, Amsterdam, The Netherlands*
[2]*University of Amsterdam, Department of Psychological Methods, Weesperplein 4, Room 218, 1018 XA, Amsterdam, The Netherlands*

The validity of a Dutch self-report inventory of dyslexia was ascertained in two samples of students. Six biographical questions, 20 general language statements and 56 specific language statements were based on dyslexia as a multi-dimensional deficit.
Dyslexia and non-dyslexia were assessed with two criteria: identification with test results (Sample 1) and classification using biographical information (both samples). Using discriminant analyses, these criteria were predicted with various groups of statements. All together, 11 discriminant functions were used to estimate classification accuracy of the inventory. In Sample 1, 15 statements predicted the test criterion with classification accuracy of 98%, and 18 statements predicted the biographical criterion with classification accuracy of 97%. In Sample 2, 16 statements predicted the biographical criterion with classification accuracy of 94%. Estimations of positive and negative predictive value were 89% and 99%.
Items of various discriminant functions were factor analysed to find characteristic difficulties of students with dyslexia, resulting in a five-factor structure in Sample 1 and a four-factor structure in Sample 2. Answer bias was investigated with measures of internal consistency reliability.
Less than 20 self-report items are sufficient to accurately classify students with and without dyslexia. This supports the usefulness of self-assessment of dyslexia as a valid alternative to diagnostic test batteries.

Keywords: dyslexia; diagnosis; adults; screening

INTRODUCTION

In the Netherlands, relatively cheap and standardized methods of diagnosing dyslexia in children have been developed, which can be assessed in schools and without the supervision of specialists (e.g. Vorst, 2006). Generally, diagnosing dyslexia in children is based on assessments of reading and/or spelling, cognitive performances (e.g. phonological), and resistance to intervention, usually in combination with the exclusion of other possible explanations such as low schooling, low intelligence or attention disorders (e.g. attention deficit hyperactivity disorder

*Correspondence to: Peter Tamboer, Weesperplein 4, Room 218, 1018 XA, Amsterdam, The Netherlands. E-mail: petertamboer@kpnmail.nl

[ADHD]). Controversy about the best diagnosing methods remains, although there is general agreement about the definition by the World Health Organization (2015) which characterizes dyslexia as a specific reading disorder characterized by a specific and significant impairment in the development of reading skills that is not solely accounted for by mental age, visual acuity problems, or inadequate schooling.

For an adult population, however, the diagnosing methods used for children may not be appropriate because in contrast with children, adults may have compensated to some extent for various difficulties accompanying dyslexia. Alternatively, various self-rating scales of dyslexia have recently been proposed (e.g. Snowling et al., 2012). However, the usefulness of self-report inventories has not fully been explored yet. Therefore, we investigated in this study the possibility of diagnosing adult dyslexia with a new self-report inventory as an alternative to test batteries in a large student sample.

In general, a problematic issue for diagnosing dyslexia both in children and in adults is that underlying aetiologies of dyslexia are still unknown, which makes it hard to interpret poor and normal performances on tests of dyslexia (Ramus & Ahissar, 2012). An additional issue for the interpretation of performances of especially adults with dyslexia is that it is unknown to what extent various coping strategies influence these performances. Adults with dyslexia who never received specialized training at school, or who even might not be aware of being dyslexic, were forced to deal with all kinds of language-related difficulties, which might differ to a large extent on an individual level. These individual differences might depend on level of intelligence, level of education, socio-economical status or any combination of these. And it is unknown how these individually different coping strategies evolve over the course of years. The main question of this study is whether these issues affect self-report questionnaires of dyslexia in the same way as they may influence test performances.

For many disorders in general, self-report inventories have proven to be reliable instruments in adult samples both for clinical diagnoses and for quantitative measurements. For example, the Beck Depression Inventory (Beck et al., 1961), one of the most widely used self-rating scales for measuring depression, was shown to have a high concurrent validity with high positive and negative predictive values (Furlanetto et al., 2005). Moreover, not only disorders but also common psychological traits can reliably be measured with self-rating scales, such as alexithymia (Vorst & Bermond, 2001) and the Big Five personality traits (Soto & John, 2009). Because self-report inventories have proven to be highly reliable and valid as a classification instrument or as a rating scale, we felt that a thorough investigation of the predictive power of a self-assessment of dyslexia is warranted.

In several previous studies, support was found for the reliability and validity of self-assessment of dyslexia. Table 1 provides an overview of the main findings. Schulte-Körne et al. (1997) found a high predictive validity of 12 general questions related to reading and spelling in a sample of reading/spelling disabled German adults. Lefly and Pennington (2000) found that a 23-item self-report of reading history had a high predictive and convergent validity in a sample of English-speaking adults with reading disability diagnosis. Wolff and Lundberg (2003) found that a 20-item self-report of symptoms of dyslexia was a more powerful predictor of Swedish adults with and without dyslexia than various phonological tests. This self-report questionnaire consisted of two scales, reading interests and other

Table 1. Overview of previous studies

Study	Number and nature of questions	Factor structure	Reliability	Sample	Criterion	Classification accuracy	Convergent validity
Schulte-Körne, Deimel and Remschmidt (1997)	12 general questions reading/spelling (two sum scores)			Elderly adults (German) 24 reading disabled 55 controls	Family genetic study	Discriminant analysis: sensitivity 71% specificity 98% Classification And Regression Tree (CART) analysis: sensitivity 87% specificity 91%	
Lefly and Pennington (2000)	23 general questions reading/spelling (5-point Likert scale)		Cronbach's alpha = .94 test–retest: .87	Elderly adults (English) 22 reading disabled 40 controls	Familial dyslexia	Discriminant analysis: sensitivity 81.8% specificity 77.5%	Correlation with reading quotient $r = .70$
Wolff and Lundberg (2003)	20 statements dyslexic symptoms (4-point scale)	Principal Components Analysis (PCA) 1. Dyslexia symptoms (14 items) 2. Reading interest (6 items)	Component: 1. alpha = .84 2. alpha = .81	Students (Swedish) 50 reading disabled/dyslexic 67 controls	Teacher rating in education centre	More powerful discriminator than phonological tests	
Mortimore and Crozier (2006)	13 general questions (yes/no) Chi-square differences			Students (English) 62 dyslexics 74 controls	Diagnosed by educational psychologist		
Snowling et al. (2012)	21 general questions (5-point or yes/no)	Conf. factor analysis 1. Reading 2. Word finding 3. Attention 4. Hyperactivity	Factors: 1. alpha = .81 2. alpha = .60 3. alpha = .81 4. alpha = .58	Adults (English) 170 family history of dyslexia 91 families of children with language difficulties 156 typical developing		Logistic regression sensitivity (poor readers): 62.5% specificity (normal readers): 95.0%	Correlations between reading and tests of spelling (.60) word reading (.51) nonword reading (.66) low correlations with between other factors and tests
Bjornsdottir et al. (2013)	23 questions (translated from Lefly & Pennington) (5-point Likert scale)	Exploratory factor analysis: 1. Dyslexia symptoms 2. Current reading 3. Memory	Cronbach's alpha = .92 test–retest .93 419 specific reading disorder: .88 679 controls: .89	Young adults (Icelandic) 419 specific reading disorder (mean 15.5, *SD* 7.5) 679 with dyslexia	Diagnosed as a child by specialist in developmental neuropsychology	ROC curve analysis: sensitivity 84.5% specificity 83.7%	

symptoms of dyslexia. Mortimore and Crozier (2006) created a 13-item questionnaire that provided a thorough overview of perceived difficulties (e.g. note taking, organization of essays and expressing ideas in writing) of people with dyslexia in British higher education. Snowling et al. (2012) designed a 15-item questionnaire that rates reading, ADHD and related skills. They found a four-factor structure (Reading, Word Finding, Attention and Hyperactivity) with the Reading Factor correlating with measured skills of spelling and decoding fluency in English-speaking adults. Bjornsdottir et al. (2013) found a three-factor structure (dyslexia symptoms, current reading and memory) in a 23-item self-report of reading history in Icelandic adults. They found that these questions quite reliably identified those with and without dyslexia according to an extern specialist. The general finding of predictive value of self-report questions is what we found as well in a previous study (Tamboer, Vorst, & Oort, 2014a), in which some self-report statements contributed to the identification of students with and without dyslexia.

The aim of this study was to further investigate reliability and validity of self-assessment of dyslexia. Although all previous studies reported relatively high validity, findings were reported based on different criteria of dyslexia or reading difficulties, and analysed with different small sets of self-report questions. Both of these differences depend on the operational definition of dyslexia, which is used. While the definition of the World Health Organization is focused on reading difficulties, other definitions are also based on, for example, phonological difficulties (e.g. Snowling et al, 1997). Furthermore, while it is still unknown what causes dyslexia, it is also unknown to what extent subtypes of dyslexia should be distinguished (Bosse et al., 2007). Many symptoms of dyslexia have been reported (e.g. Ramus & Ahissar, 2012), and especially in adult samples, it may be expected that coping strategies are based on various combinations of symptoms and differ on an individual level. It is unclear how this affects diagnoses of dyslexia in adult samples. With regard to these issues, we distinguished two main issues: the issue of criterion validity and the issue of construct validity.

Regarding criterion validity in samples of students, Mortimore and Crozier (2006) stated, '… the researcher is dependent on the criteria that are adopted by each higher education institution, and can expect that the students will have been diagnosed by a large number of professionals applying a range of different tests and procedures' (p. 239). In other words, when predictive validity of any diagnostic instrument is investigated, classification accuracy can only objectively be interpreted when an objective criterion is used. To circumvent the problem that different choices can be made for a criterion of dyslexia, we investigated two different criteria of dyslexia. One criterion was based on the previous study of ours, in which people with and without dyslexia were identified with a large battery of tests. A second criterion was based on a biographical self-report. Although we considered the first criterion to be more reliable, we hypothesized that both criteria should suffice for high criterion validity.

Regarding construct validity, Mortimore and Crozier (2006) acknowledged that dyslexic students' difficulties are not 'restricted to reading, spelling and writing, but may be experienced across the range of tasks that students encounter in higher education' (p. 236). These researchers collected various difficulties of dyslexics with only 13 items. Also in the other studies, the number of items was small, ranging from 12 to 23. The main feature of all these items is that they assess general difficulties of people with dyslexia. For instance, it is clear that the question

'Can you read quickly and easily?' (Snowling et al., 2012) assesses reading. However, reading requires several complex cognitive abilities of which it is unknown how they are related to other symptoms of dyslexia such as phonological or visual impairments.

Nevertheless, three studies (Wolff & Lundberg, 2003; Snowling et al., 2012; Bjornsdottir et al., 2013) reported factor structures in the small sets of items, although the number of factors is small (two or three) compared with the assumed number of symptoms involved in dyslexia. In order to cover as much symptoms of dyslexia as possible, we used items of the 'Communication Questionnaire' (Kramer & Vorst, 2007), which assesses various language difficulties related to dyslexia according to a facet structure. In this study, we used—besides 20 general statements comparable with the statements or questions of the previous studies—56 of these items. The main characteristic of these items is that they do not assess general language difficulties but, instead, very specific language difficulties. The idea was that the language difficulties of dyslexics are very specific and subtle, and may depend on different combinations of different symptoms of dyslexia (e.g. phonological, visual and attentional), which may be overlooked when only general statements or questions are used. We hypothesized that using a large numbers of these specific items, a factor structure might be found consisting of many factors, which would increase construct validity.

For this study, we created a new self-report inventory of dyslexia, which consists of three parts: 6 biographical questions (BQ), 20 general language statements (GLS) and 56 specific language statements (SLS). With these items, we performed prediction analyses in two samples of students and investigated classification accuracy. In the first sample, predictions were based on two criteria, one based on test results and one based on biographical information. In the second sample, only a biographical criterion was available. Furthermore, we investigated reliabilities and performed factor analyses on the sets of items that discriminated best between students with and without dyslexia.

METHODS

Participants

We used two samples of Dutch first-year psychology students. In both samples, we excluded students with missing variables, students who had a history of moderate to severe head trauma or severe general health problems, students with other disorders such as ADHD, students who were not 'completely Dutch' (we required that students were raised and went to school in the Netherlands with Dutch as their mother language) and students of whom the data showed answer patterns. Furthermore, we selected only students with ages between 17 and 25 years.

Sample 1 consisted of 418 students (104 males, mean age 19.7 [1.5] years): 67 dyslexics (20 males, mean age 20.0 [1.7] years) and 351 non-dyslexics (84 males, mean age 19.6 [1.5] years). Sample 2 consisted of 405 students (114 males, mean age 20.6 [1.4] years). Dyslexia and non-dyslexia in this sample was investigated in this study. In Sample 1, the prevalence of dyslexia may seem relatively large (67/418 = 16%). However, prevalence of dyslexia may have been underestimated to some extent, especially in a group of psychology students. In the Netherlands, psychology is considered to be a relatively easy study as compared with most other studies. Although it has not been investigated, it seems plausible that students with dyslexia prefer a relatively easy study over difficult studies.

Criteria of Dyslexia

We determined two criteria of dyslexia and non-dyslexia. A 'test criterion' was determined in order to validate the self-inventory with an independent and reliable criterion that is based on standard assessments of various symptoms of dyslexia. We used the data of 10 tests that were collected in a previous study (Tamboer, Vorst, & Oort, 2014a). These tests are described in the next paragraph. We classified all students as dyslexic or not dyslexic in a procedure consisting of two steps. First, we determined that 27 students had a formal diagnosis of dyslexia. In the Netherlands, children with dyslexia can acquire an official document that states that they are dyslexic. This document can only be provided by a specialized educational psychologist and is based on an extended testing procedure. In the Netherlands, this document is widely accepted to be a reliable diagnosis of dyslexia. In the second step, we compared the performances of this dyslexic group with the performances of the remaining students on the whole battery of 10 tests. We applied two discriminant and two logistic regression analyses on the tests. One discriminant analysis and one logistic regression analysis were based on sum scores, and one discriminant analysis and one logistic regression analysis were based on item scores. Students were categorized as dyslexic when all analyses predicted a student to be dyslexic. The 27 students with an official document were all predicted to be dyslexic according to our test battery, which validates this battery as a reliable instrument for categorizing students with dyslexia (see also, Tamboer, Vorst, & Oort, 2014a). Of the remaining students, 40 were categorized as dyslexic. Students were categorized as non-dyslexic when all analyses predicted a student to be non-dyslexic, which resulted in a group of 351 students.

A 'biographical criterion' was determined in order to establish whether it is possible to diagnose dyslexia with a self-inventory without making use of additional tests. In the present study, a biographical criterion was based on six biographical self-report questions. One question was 'Are you dyslexic?' Five other questions were self-assessments of reading, spelling and writing difficulties as a child and in the present (questions 6–10 in Appendix A). All questions had three answer categories (yes = 2, doubt = 1 and no = 0). A sum score was calculated by adding the six scores, which resulted in a score ranging from 0 to 12. Cut-off scores were determined in this study and are described in the Results section. We determined criterion groups of dyslexia and non-dyslexia using a cut-off score that was based on the lowest estimation of prevalence using the test criterion, which was 14% (see also Table 2 in the Results section). We studied the distribution of the biographical score and found that 55 students (13.2%) had a score of 5 or higher and that 75 students (17.9%) had a score of 4 or higher. We decided to use the conservative cut-off score of 5 or higher for dyslexic students (55) and 4 or lower for non-dyslexic students (363). Of the 55 dyslexic students, 51 (93%) were also identified as dyslexic with the test criterion. And of the 363 non-dyslexic students, 347 (96%) were also identified as non-dyslexic with the test criterion. Thus, 398 of 418 students (95%) were identically identified with both criteria.

Assessment of Dyslexia: A Test Battery

The test battery, used for a criterion of dyslexia, consisted of 10 tests. In the Netherlands, diagnosing test batteries usually consists of measurements of six abilities that are assumed to be impaired in people with dyslexia: reading words and/or pseudowords, spelling, phonological awareness, short-term memory, rapid naming and visual attention. A problematic issue for the present study was that some tests used in these batteries are suitable only for children but not for highly educated students. A second problematic issue was that some tests require individual testing, as is the case with rapid naming tests. We were not able to individually test more than 400 students. We therefore decided to adjust existing tests, originally designed for children, so that the tests are suitable for students and so that all tests can be assessed in large groups of students without the requirement of individual observers. In a follow-up study (in preparation), we validated the test battery of the present study with a test battery that is used in clinical settings in a small sample of students (dyslexia and non-dyslexia based on official records) that performed both test batteries. We compared predictions of both batteries and found consistent classifications in about 95% of the cases in both the dyslexic and the non-dyslexic groups of students.

All tests are explained. We did not categorize these tests into the six categories, because we assumed that some tests may require various cognitive abilities. For example, pseudoword reading is assumed to require phonological abilities, but it cannot be ruled out that abilities of attention or short-term memory are required for good performances as well.

(1) *Dutch dictation* consists of 10 sentences in the Dutch language (maximum score $10 \times 4 = 40$) and aims to measure spelling abilities. This test was a reconstruction of various existing tests for children but with a higher level of spelling difficulties. Each sentence was read aloud and could be heard through headphones. The whole sentence had to be typed into the computer. Each sentence consisted of a few words that are vulnerable to spelling errors.
(2) *English dictation* consists of 10 sentences in the English language (maximum score $10 \times 2 = 20$) and aims to measure spelling abilities. This test was a reconstruction of various existing tests for children but with a higher level of spelling difficulties. Each sentence was read aloud and could be heard through headphones. At the same time, this sentence could be read on the computer screen except two words. These words had to be retyped into the computer. Although Dutch students are familiar with the English language, some English words are well known for their vulnerability to spelling errors for Dutch people.
(3) *Missing letters* consists of 10 sentences in the Dutch language (maximum score $10 \times 2 = 20$) and aims to measure reading and spelling abilities as well as abilities of attention. This test was a reconstruction of various existing tests for children but with a higher level of spelling difficulties. Each sentence was read aloud and could be heard through headphones. For each sentence, two words were repeated, while these words were shown on the computer screen with a few letters left out of both words. The missing letters had to be typed into the computer.
(4) *Pseudowords* consists of 30 pseudowords (maximum score 30), which are non-words that sound like real words and aims to measure reading and phonological abilities and possible other abilities such as attention. Each

pseudoword was read aloud and could be heard through headphones. At the same time, this pseudoword was presented on the computer screen. It had to be decided whether the visually presented pseudoword was spelled correctly, which was the case for half of all pseudowords. Usually pseudowords are administered the other way around with participants reading the words aloud themselves. The reason for changing this was that it would be practically impossible to get all students in private sessions for this test.

(5) *Sound deletion* consists of 20 difficult Dutch words (maximum score 20) and aims to measure phonological abilities. The test was a reconstruction of various existing tests for children but with a higher level of difficulties. Each word was read aloud and could be heard through headphones. Some of these words were pronounced correctly and some words incorrectly by leaving out or adding one letter. On the computer screen, each word was presented three times, each time with a slightly different spelling with one of them being spelled correctly according to what was pronounced. Participants had to decide which of the visually presented words was heard through the headphones. For example, the existing word 'fietsenstalling'—which means bicycle shed—was read out as 'fiestenstalling'. The possible answers were 'fietsentalling', 'fiestensalling' and 'fiestenstalling'.

(6) *Spoonerisms* consists of 20 words (maximum score 20) and aims to measure phonological abilities. The test was a reconstruction of various existing tests for children but with a higher level of difficulties. A Spoonerism is a word that consists of two existing smaller words and that still consists of two small existing words when the first letters of both small words are exchanged. Each word was read aloud and could be heard through headphones. For example, the word 'kolen-schop' had to be altered into 'scholen-kop' and typed into the computer.

(7) *Incorrect spelling* consists of 40 Dutch words (maximum score 40) and aims to measure spelling and reading abilities and abilities of rapid attention. This test was designed for this study. All words were presented on a computer screen for 50 ms each. Half of the presented words were spelled correctly, and half were spelled incorrectly. The participants had to decide whether the words were spelled correctly.

(8) *Dutch–English rhyme words* consists of 40 Dutch–English word pairs (maximum score 40), presented on the computer screen with the Dutch words on the right, and aims to measure abilities of attention and phonology. This test was designed for this study. For half of the word pairs, the nouns of the Dutch and the English word resembled each other visually but not aurally. For the other half of the word pairs, the nouns of the Dutch and the English word resembled each other aurally.

(9) *Letter order* requires the ability to read words as a whole and consisted of 20 Dutch sentences (maximum score $20 \times 2 = 40$; time limit of 5 min) and aims to measure abilities of whole word reading. This test was designed for this study. Theoretical hypotheses about reading words as a whole are described in the *dual route model of reading* (for an extended description, see De Groot et al., 1994). The idea for this test is based on a well-known text: 'Aoccdrnig to rscheearch at Cmabrigde uinervtisy, it deosn't mttaer waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteres are at the rghit pclae. The rset can be a tatol mses and you can sitll raed it wouthit a porbelm. Tihs is bcuseae we do not raed ervey lteter by it slef

but the wrod as a wlohe'. We created Dutch sentences based on the same principle; the order of letters in the words was changed, apart from the first and last letters. The sentences became more difficult towards the end of the test. The sentences had to be typed in with all words correctly spelled.

(10) The *short-term memory test* aims to measure the capacity of short-term memory and was based on various existing tests of digit span capacity. The digit span is the number of digits a person can retain and recall. There are four subtests: numbers and letters, both forward and backward. And each subtest consists of 24 series: 6 of 4, 6 of 5, 6 of 6 and 6 of 7 items for the subtests numbers and letters forward, and 6 of 3, 6 of 4, 6 of 5 and 6 of 6 items for the subtests numbers and letters backward. The numbers and letters are presented one by one, for 1 s each on a computer screen. The participants have to retype these numbers and letters after the last one of a series has been presented. About half of all series consist of some typical difficulties for dyslexic students, either phonological or visual or both. For example, a typical phonological confusion is between the numbers seven and nine, which resemble each other phonologically in Dutch (zeven/negen). Typical visual confusions are between the numbers six and nine and the letters [m] and [w]. The letters [p], [d] and [b] resemble each other phonologically as well as visually.

Procedure

All data of the first sample were collected at the University of Amsterdam in 2009 and 2010 during five sessions of 3 h each, in which tests and questionnaires of various studies were administered. These sessions took place on midweek evenings with 1 or 2 weeks between each session. All students were obligated to participate because these sessions were part of the first-year study programme. The students were informed about the general nature of the tests and the questionnaires in advance according to a standard protocol. Afterwards, the students received a more detailed debriefing. Anonymity was guaranteed by the standard protocol of the University of Amsterdam. The students had up to 3 weeks after their debriefing to request that their test results were not used. The data of the second sample were collected in 2011 in one session and under the same conditions.

Self-Report Inventory

The self-report inventory of this study consisted of three parts: 6 BQ, 20 GLS and 56 SLS. See Appendix A. There are two differences regarding the two samples. In the first sample, the three types of questions were collected in separate and larger questionnaires consisting of more items, while in the second sample, a selection of all questions was presented in one questionnaire. And the students of the first sample did not know that the questions were aiming to measure dyslexic difficulties (except for the BQ), while in the questionnaire for the second sample, it was mentioned that at least some of the questions were aiming to measure dyslexia.

Biographical Self-Report Questions

Biographical information regarding dyslexia is assessed with six questions. The first question is a direct self-report question: 'Are you dyslexic?' Five questions request to assess to what extent performances of spelling, reading and writing were poor in the past and in the present.

General Language Statements

In our first study (Tamboer, Vorst, & Oort, 2014a), we created a questionnaire that consisted of 60 general statements about reading, speaking, writing, mental representations, memory and foreign languages. We selected the 20 best discriminating statements based on group differences. For each statement, there are seven response categories (7-point Likert scale).

Specific Language Statements

In our first study, we assessed the *Questionnaire Communication* (Kramer & Vorst, 2007) that consists of 140 statements about specific language-related difficulties. The questionnaire was designed according to a $7 \times 5 \times 4$ facet design with seven response categories (7-point Likert scale). One dimension distinguishes between seven different aspects of how language is used in daily life or at school and universities: reading, writing, speaking, listening, copying, taking a dictation and reading aloud. A second dimension distinguishes between five different levels of how language can be represented: by sounds, letters, words, sentences or text. A third dimension distinguishes between four different difficulties that accompany dyslexic adults: skipping (forgetting), adding, changing and exchanging. For instance, dyslexic students may skip parts of sentences when reading, exchange letters when writing, change words when speaking or forget parts of texts when taking a dictation.

Support for impairments related to all aspects of the first and second dimensions can be found in the literature. The difficulties of the third dimension need some clarification. They are related to some kind of attention or organization skills or concentration, which were found to be impaired in people with dyslexia in higher education (Mortimore & Crozier, 2006). These researchers presented an overview of difficulties that very much resemble the difficulties represented by the three dimensions of the *Questionnaire Communication*. They distinguished 13 items representing reading, reading speed, spelling, note taking, organizing essays, general organization, time keeping, expressing ideas orally, expressing ideas in writing, handwriting, concentration, remembering facts and listening. All difficulties were reported significantly more often by those with dyslexia than by those without dyslexia.

For the present study, we selected 56 items with a high predictive value while preserving the original facet design as much as possible.

Tests of Intelligence

For being able to analyse convergent and divergent validity of the questionnaire, we needed additional information about performances on tests of intelligence. Available were the data of six cognitive tests that were based on Guilford's Structure of Intellect Model (Vocabulary, Verbal analogies, Conclusions, Numeric progressions, Speed of Calculation and Hidden figures) and for general (non-verbal) intelligence (Advanced Progressive Matrices Set 2, Raven, Court, & Raven, 1979).

Table 2. Predictions in Sample 1 with test criterion

Discriminant function 1 was generated with:			
Potential predictors:			6 BQ
Criterion of dyslexia:			Test battery
Criterion groups:			67 dyslexics and 351 non-dyslexics
Results:			
Number of selected predictors in discriminant function:			5 BQ
Percentage correctly classified of 418 students:			94%
Estimated prevalence of dyslexia:			15%
Classification of 67 dyslexic and 351 non-dyslexia students:			
	Predicted as:		
	D	ND	Total
D	50 (78%)	17 (22%)	67
ND	6 (2%)	345 (98%)	351
	56 (15%)	362 (85%)	418
Discriminant function 2 was generated with:			
Potential predictors:	20 GLS		
Criterion of dyslexia:	Test battery		
Criterion groups:	67 dyslexics and 351 non-dyslexics		
Results:			
Number of selected predictors in discriminant function:			11 GLS
Percentage correctly classified of 418 students:			88%
Estimated prevalence of dyslexia:			21%
Classification of 67 dyslexic and 351 non-dyslexic students:			
	Predicted as:		
	D	ND	Total
D	53 (79%)	21 (18%)	67
ND	35 (10%)	31 (90%)	351
	88 (21%)	330 (79%)	418
Discriminant function 3 was generated with:			
Potential predictors:	56 SLS		
Criterion of dyslexia:	Test battery		
Criterion groups:	67 dyslexics and 351 non-dyslexics		
Results:			
Number of selected predictors in discriminant function:			12 SLS
Percentage correctly classified of 418 students:			94%
Estimated prevalence of dyslexia:			19%
Classification of 67 dyslexics and 351 non-dyslexics:			
	Predicted as:		
	D	ND	Total
D	61 (91%)	6 (9%)	67
ND	17 (5%)	334 (95%)	351
	78 (19%)	340 (81%)	418

(*Continues*)

Table 2. (Continued)

Discriminant function 4 was generated with:			
Potential predictors:			20 GLS + 56 SLS
Criterion of dyslexia:			Test battery
Criterion groups:			67 dyslexics and 351 non-dyslexics
Results:			
Number of selected predictors in discriminant function:			6 GLS + 11 SLS
Percentage correctly classified of 418 students:			97%
Estimated prevalence of dyslexia:			17%
Classification of 67 dyslexic and 351 non-dyslexic students:			
	Predicted as:		
	D	ND	Total
D	64 (96%)	3 (4%)	67
ND	8 (2%)	343 (98%)	351
	72 (17%)	346 (83%)	418
Discriminant function 5 was generated with:			
Potential predictors:			6 BQ + 20 GLS + 56 SLS
Criterion of dyslexia:			Test battery
Criterion groups:			67 dyslexics and 351 non-dyslexics
Results:			
Number of selected predictors in discriminant function:			4 BQ + 2 GLS + 9 SLS
Percentage correctly classified of 418 students:			98%
Estimated prevalence of dyslexia:			14%
Classification of 67 dyslexic and 351 non-dyslexic students:			
	Predicted as:		
	D	ND	Total
D	58 (87%)	9 (13%)	67
ND	1 (0%)	350 (100%)	351
	59 (14%)	359 (86%)	418

D, dyslexic; ND, non-dyslexic; BQ,biographical questions; GLS, general language statements; SLS, specific language statements.

Characteristics of Samples 1 and 2

In Sample 1, the three parts of the self-report inventory were administered on three different days. The part consisting of the GLS and the part consisting of the SLS were both administered without the participants knowing that the statements were aiming to assess dyslexia. Furthermore, data of dyslexia tests and intelligence tests were available. For analyses of criterion validity, criterion groups were determined with two criteria of dyslexia, a test criterion and a biographical criterion.

In Sample 2, the three parts of the self-report inventory were administered together. In contrast to the participants of Sample 1, the participants of Sample 2 were aware that the statements were aiming to assess dyslexia. Data of dyslexia tests and intelligence tests were not available for Sample 2. For analyses of criterion validity, criterion groups were determined with only a biographical criterion.

Analyses

We investigated criterion and construct validity in five steps. In the first step, we looked for differences between dyslexic and non-dyslexic students on sum scores of the 20 GLS, the 56 SLS and all statements together. Furthermore, we looked for differences between students with and without dyslexia on each individual item.

In the second step, we performed predictions with five different sets of potential predictors: 6 BQ, 20 GLS, 56 SLS, the 76 GLS and SLS together, and all 82 items together. In Sample 1, we investigated classification accuracy of these predictions with a test criterion and a biographical criterion. In Sample 2, we investigated classification accuracy with a biographical criterion. Next, we validated the best discriminant functions of Sample 1 in Sample 2. All predictions were carried out with discriminant analysis (stepwise method) in SPSS. We also investigated whether results differed with or without the possibility of cross-validation, the leave-one-out procedure in SPSS.

In the third step, we compared the results of the prediction analyses with measures of internal consistency. We calculated Cronbach's alpha, the Spearman–Brown coefficient and the Guttman Split-Half coefficient of all 76 items and of the 20 GLS and the 56 SLS separately. We did this in each sample for the whole group, the dyslexic and non-dyslexic groups. We also calculated the mean of all items' means and the mean of all items' variance.

In the fourth step, we used the items that were selected for various discriminant functions for exploratory factor analyses. In another previous study (Tamboer, Vorst, & Oort, 2014b), we found a five-factor structure in a confirmatory factor model with five factors representing variance in the 10 tests. We tested whether this five-factor structure (spelling, phonology, short-term memory, rhyme/confusion and whole word processing/complexity) could be replicated.

In the fifth step, we investigated convergent and divergent validity with correlations between sum scores of the self-report inventory and sum scores of the tests and intelligence measures.

RESULTS AND CONCLUSIONS

Group Differences

We calculated group differences between the dyslexic (67) and non-dyslexic (351) groups of Sample 1 (according to a test criterion). First, we calculated group means and standard deviations of the sum scores of the 20 GLS, the 56 SLS and all items together. Students with dyslexia had lower scores than students without dyslexia on all three sets of items ($p < 0.001$). Second, we calculated group means and standard deviations of all individual items. Students with dyslexia had lower scores than students without dyslexia on all items ($p < 0.05$), except for seven items of the 20 GLS (nos. 1, 5, 7, 11, 15, 17 and 20) and two items of the 76 SLS (nos. 23 and 26). The items are presented in Appendix A.

Criterion Validity

Predictions in Sample 1 with test criterion

In Sample 1, we first investigated criterion validity with a test criterion and five sets of items as potential predictors. With each set, a discriminant function (DF) was determined, which maximally separated the test-criterion groups of 67 students with dyslexia and 351 students without dyslexia. DF1 was acquired with 6 BQ as potential predictors, DF2 was acquired with 20 GLS as potential predictors, DF3 was acquired with 56 SLS as potential predictors, DF4 was acquired with

76 GLS and SLS together as potential predictors, and DF5 was acquired with all 82 items together as potential predictors. All discriminant functions are presented in Appendix B. Table 2 presents classification accuracies of these discriminant functions in five cross-tables. For each DF, specific details are presented. First, it is specified how the DF was generated: the number of potential predictors, the criterion that was used and the criterion groups. Second, the results of the analysis are specified: the number of selected predictors in the DF, the percentage correctly classified students with and without dyslexia and the estimated prevalence of dyslexia. Third, classification accuracy is presented in a cross-table. The left percentage above represents sensitivity (number of correctly classified students with dyslexia), and the right percentage below represents specificity (number of correctly classified students without dyslexia). Completely below, the total percentages of predicted students with and without dyslexia are presented.

DF4 and DF5 are the best functions for predicting dyslexia and non-dyslexia according to a test criterion. DF4 (6 GLS + 11 SLS) resulted in classification accuracy of 97%, with sensitivity of 96% and specificity of 98%, and with an estimated prevalence of dyslexia of 17%. DF5 (4 BQ + 2 GLS + 9 SLS) resulted in classification accuracy of 98%, with sensitivity of 87% and specificity of 100%, and with an estimated prevalence of dyslexia of 14%. Lower classification accuracies were found for DF3 (12 SLS, 94%) and DF2 (11 GLS, 88%). DF1 (5 BQ) resulted in classification accuracy of 94%. The importance of this last finding is that a test criterion is relatively consistent with a biographical criterion.

We conclude that using the combination of GLS and SLS (and BQ) as predictors results in a classification of students with and without dyslexia which is highly consistent with a classification using test results. An important remark is that we only reported classification results in the tables that were acquired without using a cross-validation procedure (leave-one-out). We did this because we want to use the same discriminant formulas that were acquired in Sample 1 in Sample 2, thus validating the results of Sample 1 to Sample 2. Nevertheless, we also repeated the best predictions, DF4 and DF5, using a cross-validation procedure (leave-one-out). For DF4, we found that two more dyslexic students and two more non-dyslexic students were classified incorrectly, leading to classification accuracy of 96%, a reduction of 1%. For DF5, we found that one more dyslexic student and one more non-dyslexic student were classified incorrectly, leading to classification accuracy of 97%, a reduction of 1%. These differences are so small that we conclude that validity is not threatened when no cross-validation is applied.

Predictions in Sample 1 with biographical criterion

An extended criterion, such as the test criterion in this study, is not always available. Therefore, we wanted to investigate whether an easy to acquire biographical criterion can be used for a reliable classification of dyslexia and non-dyslexia. We investigated criterion validity with this biographical criterion in the same way as we did with the test criterion. Three sets of items were used as potential predictors. With each set, a discriminant function was determined, which maximally separated the 55 dyslexic and 363 non-dyslexic students according to the biographical criterion. DF6 was acquired with 20 GLS as potential predictors, DF7 was acquired with 56 SLS as potential predictors and DF8 was acquired with 76 GLS and SLS together as potential predictors. All discriminant functions are presented in Appendix B. We investigated classification accuracy in two different ways.

Table 3. Predictions in Sample 1 with biographical criterion (classification of biographical criterion groups)

Discriminant function 6 was generated with:	
Potential predictors:	20 GLS
Criterion of dyslexia:	Biographical criterion
Criterion groups:	55 dyslexics and 363 non-dyslexics
Results:	
Number of selected predictors in discriminant function:	12 GLS
Percentage correctly classified of 418 students:	86%
Estimated prevalence of dyslexia:	22%

Classification of 55 dyslexic and 363 non-dyslexic students:

	Predicted as: D	ND	Total
D	43 (78%)	12 (22%)	55
ND	47 (13%)	316 (87%)	363
	90 (22%)	328 (78%)	418

Discriminant function 7 was generated with:	
Potential predictors:	56 SLS
Criterion of dyslexia:	Biographical criterion
Criterion groups:	55 dyslexics and 363 non-dyslexics
Results:	
Number of selected predictors in discriminant function:	16 SLS
Percentage correctly classified of 418 students:	94%
Estimated prevalence of dyslexia:	17%

Classification of 55 dyslexic and 363 non-dyslexic students:

	Predicted as: D	ND	Total
D	50 (91%)	5 (9%)	55
ND	20 (6%)	343 (94%)	363
	70 (17%)	348 (83%)	418

Discriminant function 8 was generated with:	
Potential predictors:	20 GLS + 56 SLS
Criterion of dyslexia:	Biographical criterion
Criterion groups:	55 dyslexics and 363 non-dyslexics
Results:	
Number of selected predictors in discriminant function:	3 GLS + 15 SLS
Percentage correctly classified of 418 students:	97%
Estimated prevalence of dyslexia:	15%

Classification of 55 dyslexic and 363 non-dyslexic students:

	Predicted as: D	ND	Total
D	53 (96%)	2 (4%)	55
ND	11 (3%)	352 (97%)	363
	64 (15%)	354 (85%)	418

D, dyslexic; ND, non-dyslexic; GLS, general language statements; SLS = specific language statements.

Table 4. Predictions in Sample 1 with biographical criterion (classification of test-criterion groups)

Discriminant function 6 was generated with:	
Potential predictors:	20 GLS
Criterion of dyslexia:	Biographical criterion
Criterion groups:	55 dyslexics and 363 non-dyslexics
Results:	
Number of selected predictors in discriminant function:	12 GLS
Percentage correctly classified of 418 students:	89%
Estimated prevalence of dyslexia:	22%

Classification of 67 dyslexic and 351 non-dyslexic students:

	Predicted as: D	ND	Total
D	55 (82%)	12 (18%)	67
ND	35 (10%)	316 (90%)	351
	90 (22%)	328 (78%)	418

Discriminant function 7 was generated with:	
Potential predictors:	56 SLS
Criterion of dyslexia:	Biographical criterion
Criterion groups:	55 dyslexics and 363 non-dyslexics
Results:	
Number of selected predictors in discriminant function:	16 SLS
Percentage correctly classified of 418 students:	94%
Estimated prevalence of dyslexia:	17%

Classification of 67 dyslexic and 351 non-dyslexic students:

	Predicted as: D	ND	Total
D	57 (85%)	10 (15%)	67
	13 (4%)	338 (96%)	351
ND	70 (17%)	348 (83%)	418

Discriminant function 8 was generated with:	
Potential predictors:	20 GLS + 56 SLS
Criterion of dyslexia:	Biographical criterion
Criterion groups:	55 dyslexics and 363 non-dyslexics
Results:	
Number of selected predictors in discriminant function:	3 GLS + 15 SLS
Percentage correctly classified of 418 students:	95%
Estimated prevalence of dyslexia:	15%

Classification of 67 dyslexic and 351 non-dyslexic students:

	Predicted as: D	ND	Total
D	56 (84%)	11 (16%)	67
	8 (2%)	343 (98%)	351
ND	64 (15%)	354 (85%)	418

D, dyslexic; ND, non-dyslexic; GLS, general language statements; SLS, specific language statements.

First, we examined classification accuracy in the dyslexic and non-dyslexic groups as they were selected by the biographical criterion. Table 3 presents classification accuracies in three cross-tables. For each discriminant function, specific details are presented in the same way as in Table 2. DF8 is most consistent with the classification based on a biographical criterion. DF8 (3 GLS + 15 SLS) resulted in classification accuracy of 97%, with sensitivity of 96% and specificity of 97%, and with an estimated prevalence of dyslexia of 15%. When we repeated the analysis with a cross-validation procedure (leave-one-out), we found slightly lower classification accuracy of 95%.

Second, we repeated these analyses but now using dyslexic and non-dyslexic students as they were selected by the test criterion. In other words, when we assume that the test criterion is more reliable than the biographical criterion, we should establish whether the dyslexic and non-dyslexic students according to the test criterion are correctly classified. This should be considered as a better test of validity than looking how the dyslexic and non-dyslexic students according to the biographical criterion were classified, because the predictions are completely independent from the criterion. Nevertheless, that 95% of the 418 students were identified identically with both criteria means that the use of a biographical criterion for prediction analyses is probably justified. Table 4 presents classification accuracies in three cross-tables. For each discriminant function, specific details are presented in the same way as in Tables 2 and 3. DF8 is the best function for predicting dyslexia and non-dyslexia according to a test criterion. DF8 (3 GLS + 15 SLS) resulted in classification accuracy of 95%, with sensitivity of 84% and specificity of 98%, and with an estimated prevalence of dyslexia of 15%.

Findings and conclusions (Sample 1)

With the analyses using a test criterion, we found that DF4 (6 GLS + 11 SLS) and DF5 (4 BQ + 2 GLS + 9 SLS) resulted in highest classification accuracy (97% and 98%) with an estimation of prevalence of dyslexia of 17% and 14%. Also with the biographical criterion, we found that using all items as potential predictors, thus DF8 (3 GLS + 15 SLS), resulted in the highest classification accuracy (98%) with an estimation of prevalence of dyslexia of 15%.

We considered the test criterion to represent a more reliable classification of dyslexic and non-dyslexic students than the biographical criterion. Thus, the importance of Table 4 is that it shows that—using a biographical criterion—the self-report inventory reliably identifies most of the 'true' dyslexic and non-dyslexic students. On the other hand, when a test criterion would not have been available, the importance of Table 3 is that it shows that the self-report inventory is highly consistent with a biographical criterion.

We found that both a biographical criterion and a test criterion resulted in high classification accuracies, especially when all items are used as potential predictors. Specific differences can be found when we compare Table 2 with Table 4. In other words, comparing these tables means comparing classification accuracy in the same dyslexic and non-dyslexic groups, but with two different criteria. Regarding the GLS as potential predictors, we should compare the results of DF2 with the results of DF6. There are hardly any differences. Regarding the SLS as potential predictors, we should compare the results of DF3 with the results of DF7. The only difference here is that sensitivity is higher when the test criterion is used. Regarding all items as potential predictors, we should compare the results of DF4 with the results of DF8. Also here, the main difference is that sensitivity is higher when the test criterion is used. The conclusion here is that validity (sensitivity) is higher when a test criterion is used. In other words, with the inventory, most of the non-dyslexic students (98%) are always correctly identified, but dyslexic students are better identified when a test criterion is available than when only a biographical criterion is available.

Table 5. Predictions in Sample 2 with biographical criterion

Discriminant function 9 was generated with:

Potential predictors:	20 GLS
Criterion of dyslexia:	biographical criterion
Criterion groups:	52 dyslexics and 353 non-dyslexics

Results:

Number of selected predictors in discriminant function:	8 GLS
Percentage correctly classified of 405 students:	88%
Estimated prevalence of dyslexia:	22%

Classification of 52 dyslexic and 353 non-dyslexic students:

	Predicted as:		
	D	ND	Total
D	46 (88%)	6 (12%)	52
	42 (12%)	311 (88%)	353
ND	88 (22%)	317 (78%)	405

Discriminant function 10 was generated with:

Potential predictors:	56 SLS
Criterion of dyslexia:	Biographical criterion
Criterion groups:	52 dyslexics and 353 non-dyslexics

Results:

Number of selected predictors in discriminant function:	10 SLS
Percentage correctly classified of 405 students:	90%
Estimated prevalence of dyslexia:	19%

Classification of 52 dyslexic and 353 non-dyslexic students:

	Predicted as:		
	D	ND	Total
D	45 (87%)	7 (13%)	52
	33 (9%)	320 (91%)	353
ND	78 (19%)	327 (81%)	405

Discriminant function 11 was generated with:

Potential predictors:	20 GLS + 56 SLS
Criterion of dyslexia:	Biographical criterion
Criterion groups:	52 dyslexics and 353 non-dyslexics

Results:

Number of selected predictors in discriminant function:	5 GLS + 11 SLS
Percentage correctly classified of 405 students:	94%
Estimated prevalence of dyslexia:	17%

Classification of 52 dyslexics and 353 non-dyslexics:

	Predicted as		
	D	ND	Total
D	48 (92%)	4 (8%)	52
	21 (6%)	332 (94%)	353
ND	69 (17%)	336 (83%)	405

D, dyslexic; ND, non-dyslexic; GLS, general language statements; SLS, specific language statements.

Another way to interpret the results is to consider positive predictive value (the proportion of true dyslexic students of the total number of predicted dyslexic students) and negative predictive value (the proportion of true non-dyslexic students of the total number of predicted non-dyslexic students). We compared DF4 with DF8 (GLS and SLS as potential predictors). DF4 (test criterion) resulted in a positive predictive value of (64 / 72 =) 89% and a negative predictive value of (343 / 346 =) 99%. DF8 (biographical criterion) resulted in a positive predictive value of (56 / 64 =) 88% and a negative predictive value of (343 / 354 =) 97%. Thus, when the inventory is assessed and only a biographical criterion is available, one can conclude that of the classified dyslexic students, 88% is really with dyslexia and that of the classified dyslexic students, 97% is really without dyslexia.

Summarizing, there are four conclusions. (1) Highest validity is acquired when all items are used as potential predictors. (2) Both a test criterion and a biographical criterion can be used for reliable classifications of people with and without dyslexia, although sensitivity is higher using a test criterion. (3) With both criteria, negative predictive value is higher than positive predictive value. (4) Based on the most reliable discriminant functions, estimations of prevalence in the sample of this study vary between 14% and 17%.

Table 6. Predictions in Sample 2 by validating discriminant functions from Sample 1

Discriminant function 4:	
Potential predictors:	20 GLS + 56 SLS
Criterion of dyslexia:	Test battery in Sample 1
Criterion groups:	67 dyslexics and 351 non-dyslexics of Sample 1
Number of selected predictors in discriminant function:	6 GLS + 11 SLS
Results:	
Percentage correctly classified of 405 students:	85%
Classification of 69 dyslexic and 336 non-dyslexic students:	

	Predicted as: D	ND	Total
D	53 (77%)	16 (23%)	69
ND	45 (13%)	291 (87%)	336

Discriminant function 5:	
Potential predictors:	6 BQ + 20 GLS + 56 SLS
Criterion of dyslexia:	Test battery in Sample 1
Criterion groups:	67 dyslexics and 351 non-dyslexics of Sample 1
Number of selected predictors in discriminant function:	4 BQ + 2 GLS + 9 SLS
Results:	
Percentage correctly classified of 405 students:	80%
Classification of 69 dyslexic and 336 non-dyslexic students:	

	Predicted as: D	ND	Total
D	57 (83%)	12 (17%)	69
ND	70 (21%)	266 (79%)	336

D, dyslexic; ND = non-dyslexic; BQ, biographical questions; GLS, general language statements; SLS, specific language statements.

Predictions in Sample 2 with biographical criterion

In Sample 2, a test criterion was not available. A biographical criterion was determined in the same way as in Sample 1. In Sample 1, 55 students (13.2%) had a score of 5 or higher and 363 students (86.8%) had a score of 4 or lower. In Sample 2, 52 students (12.8%) had a score of 5 or higher and 353 students (87.2%) had a score of 4 or lower. The percentages are about the same in both samples. Therefore, we used these groups as biographical criterion groups.

Three sets of items were used as potential predictors. With each set, a discriminant function was determined, which maximally separated the 52 students with dyslexia and 353 students without dyslexia according to the biographical criterion. DF9 was acquired with 20 GLS as potential predictors, DF10 was acquired with 56 SLS as potential predictors and DF11 was acquired with 76 GLS and SLS together

Table 7. Internal consistency reliabilities (20 GLS, 56 SLS and 76 GLS + SLS)

	Cronbach's alpha	Spearman–Brown	Guttman Split-Half	Item mean (mean)	Item *SD* (mean)
20 GLS					
Sample 1					
Whole group ($n = 418$)	0.50	0.62	0.62	4.93	2.42
D test criterion ($n = 67$)	0.46	0.54	0.54	4.41	3.13
ND test criterion ($n = 351$)	0.34	0.50	0.50	5.03	2.20
D biographical criterion ($n = 55$)	0.44	0.53	0.51	4.37	3.22
ND biographical criterion ($n = 363$)	0.37	0.52	0.52	5.01	2.22
Sample 2					
Whole group ($n = 405$)	0.83	0.81	0.80	4.91	2.65
D biographical criterion ($n = 52$)	0.70	0.71	0.70	3.77	3.35
ND biographical criterion ($n = 353$)	0.78	0.72	0.72	5.08	2.29
56 SLS					
Sample 1					
Whole group ($n = 418$)	0.97	0.94	0.93	5.42	1.63
D test criterion ($n = 67$)	0.93	0.87	0.87	4.58	1.93
ND test criterion ($n = 351$)	0.96	0.92	0.92	5.58	1.37
D biographical criterion ($n = 55$)	0.93	0.87	0.87	4.56	1.98
ND biographical criterion ($n = 363$)	0.96	0.93	0.92	5.55	1.41
Sample 2					
Whole group ($n = 405$)	0.97	0.93	0.92	5.51	2.15
D biographical criterion ($n = 52$)	0.95	0.86	0.85	4.26	2.78
ND biographical criterion ($n = 353$)	0.96	0.91	0.90	5.69	1.75
76 GLS + SLS					
Sample 1					
Whole group ($n = 418$)	0.95	0.89	0.86	5.29	1.84
D test criterion ($n = 67$)	0.91	0.75	0.74	4.54	2.25
ND test criterion ($n = 351$)	0.94	0.87	0.82	5.43	1.59
D biographical criterion ($n = 55$)	0.90	0.71	0.69	4.51	2.31
D biographical criterion ($n = 363$)	0.94	0.88	0.84	5.41	1.62
Sample 2					
Whole group ($n = 405$)	0.97	0.94	0.94	5.35	2.28
D biographical criterion ($n = 52$)	0.95	0.88	0.87	4.13	2.93
D biographical criterion ($n = 353$)	0.96	0.92	0.91	5.53	1.89

D, dyslexic; ND, non-dyslexic; GLS, general language statements; SLS, specific language statements.

as potential predictors. We investigated classification accuracy with these discriminant functions. Table 5 presents all cross-tables. All discriminant functions are presented in Appendix B. For each discriminant function, specific details are presented in the same way as in Tables 2, 3 and 4.

Table 8. Exploratory factor analysis: Number of components and amount of variance explained

	Sample	*N* components	Percentage variance explained
DF3 (12 SLS, selected with test criterion):	1	2	53%
	2	2	57%
DF4 (6 GLS + 11 SLS, selected with test criterion):	1	5	60%
	2	3	54%
DF5 (4 BQ + 2 GLS + 9 SLS, selected with test criterion):	1	2	53%
	2	2	51%
DF7 (16 SLS, selected with biographical criterion):	1	3	56%
DF8 (3 GLS and 15 SLS, selected with biographical criterion):	1	4	54%
DF10 (10 SLS, selected with biographical criterion):	2	1	42%
DF11 (5 GLS and 11 SLS, selected with biographical criterion):	2	4	57%

DF, discriminant function; GLS, general language statements; SLS, specific language statements.

Table 9. Correlations between sum scores of the 56 SLS and test scores ($N = 480$)

	Spelling	Phonology	Short-term memory	Rhyme/ Confusion	Whole word processing complexity
Language-related tests					
Dutch dictation	0.43**	0.22**	0.30**	0.21**	0.31**
English dictation	0.33**	0.20**	0.24**	0.22**	0.28**
Missing letters	0.36**	0.25**	0.31**	0.26**	0.32**
Pseudowords	0.19**	0.18**	0.21**	0.13**	0.17**
Sound deletion	0.11*	0.11*	0.10*	0.05	0.10*
Spoonerisms	0.21**	0.16**	0.21**	0.14**	0.18**z
Incorrect spelling	0.41**	0.20**	0.30**	0.26**	0.32**
Dutch–English rhyme words	0.34**	0.21**	0.33**	0.25**	0.32**
Short-term memory numbers	0.35**	0.28**	0.33**	0.32**	0.37**
Short-term memory letters	0.35**	0.26**	0.32**	0.31**	0.36**
General cognitive tests					
Vocabulary	0.23**	0.14**	0.23**	0.15**	0.24**
Conclusions	0.05	−0.03	0.10*	0.07	0.08
Numeric progressions	0.09	0.05	0.10*	0.04	0.10*
Speed of calculation	0.20**	0.09*	0.16**	0.12**	0.16**
Verbal analogies	0.19**	0.11*	0.17**	0.15**	0.18**
Hidden figures	0.14**	0.07	0.14**	0.13**	0.17**
Raven progressive matrices	0.05	−0.02	0.08	0.05	0.08

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

DF11 is most consistent with the classification based on a biographical criterion. DF11 (5 GLS + 11 SLS) resulted in classification accuracy of 94%, with sensitivity of 92% and specificity of 94%, and with an estimated prevalence of dyslexia of 17%. When we repeated the analysis with a cross-validation procedure (leave-one-out), we found slightly lower classification accuracy of 91%. We concluded that the self-report inventory reliably predicts biographical criterion groups of dyslexia and non-dyslexia (94%, cross-validated 91%). However, without an alternative criterion such as a test criterion, we cannot know whether the false positives and false negatives represent false classifications of the inventory or false classifications of the criterion. It is therefore not meaningful to calculate positive predictive value and negative predictive value.

Predictions in Sample 2 by validating discriminant functions from Sample 1

Although a test criterion was not available in Sample 2, we could use the discriminant functions that were acquired in Sample 1 with a test criterion as an extern criterion in Sample 2. The best predictions in Sample 1 with the test criterion were derived with DF4 (20 GLS and 56 SLS as potential predictors) and DF5 (6 BQ, 20 GLS and 56 SLS as potential predictors). We validated these two discriminant functions in Sample 2. Thus, in Sample 1, we acquired two discriminant functions that will be tested on new cases, the students of Sample 2. Summarizing, we wanted to know whether the best classification possible with the inventory was consistent with an external criterion. The best classification in Sample 2 was the one derived with DF11 (5 GLS + 11 SLS), resulting in 69 students with dyslexia and 336 students without dyslexia.

Table 6 presents two cross-tables. For each discriminant function, specific details are presented in the same way as in the previous tables. We found a consistency of 85% between the classification with a biographical criterion and the classification with a discriminant function (DF4), which was acquired with Sample 1. There are two interpretations of this result. One is that it is shown that DF4 can be applied in a new sample and that the reliability of this discriminant function is relatively high (85%). A second interpretation is that the self-report inventory results in a relatively reliable classification of dyslexic and non-dyslexic students without the necessity of using an external criterion.

CONCLUSIONS

From the three analyses, we can summarize the following conclusions. Regarding the different criteria of dyslexia, we conclude that both a test criterion and a biographical criterion can be used for reliable prediction analyses, although validity of a test criterion was found to be higher than validity of a biographical criterion, especially regarding sensitivity. Regarding the different sets of predictors, we conclude that the combined set of GLS and SLS provided the highest classification accuracy and that not more than 20 items are needed for a reliable classification of people with and without dyslexia. Reliability of these sets was relatively successfully validated in Sample 2. However, two things are not completely clarified yet.

First, it remains unclear what the best criterion of dyslexia is. We speculated that a test criterion is more reliable than a biographical criterion. However, this cannot be proven from the results of this study. Nevertheless, in a new sample, we found that classifications with both criteria were consistent in 85% of all students. We conclude that criterion validity of the self-report inventory is relatively high, and how high depending on the interpretation of criteria but at least 85%.

Second, the relative success of the different sets of predictors might depend on the differences between the two samples. In Sample 1, the GLS and the SLS were administered on separate days and without mentioning that dyslexia was assessed. In Sample 2, the GLS and the SLS were administered in one questionnaire, while the participants were aware that the statements were aiming to assess dyslexia. This might have resulted in somewhat lower classification accuracy of 85% by validation than classification accuracies acquired within one sample. A possibility is that answer bias may have differed between the samples and between the GLS and SLS.

Internal Consistency Reliability

In previous studies (Table 1), mainly general statements were used. It is furthermore difficult to establish whether and how answer bias may have affected self-report questions or statements of dyslexia. Four studies reported measures of internal consistencies, which were all between 0.80 and 0.94, except a factor 'word finding' (0.60) in the study of Snowling et al. (2012). We found in our study that the combination of general and specific statements resulted in the highest criterion validity but that specific statements have more predictive power than general statements. In the previous paragraph, we suggested that differences between the two samples may have resulted in a decrease of classification accuracy, such as different patterns of answer bias. We further investigated this issue by analysing internal consistency reliability of the separate sets of potential predictors in the two samples.

We investigated internal consistency reliability with three separate measures: Cronbach's alpha, the Spearman–Brown coefficient and Guttman's Split-Half coefficient. These measures were calculated for the three sets of potential predictors, for Samples 1 and 2 separately, and for the dyslexic and non-dyslexic students separately and together. Furthermore, we determined the means of item mean and item standard deviation. We had three expectations. First, we expected higher values resulting from answer bias in Sample 2, because the purpose of the inventory was known in this sample. Second, we expected higher values for the SLS than for the GLS, because for the SLS, we carefully explored symptoms of dyslexia, while it is unknown whether GLS represent only difficulties that are characteristic for dyslexia. (This suggestion is supported by the fact that 74 of 76 SLS showed group differences but that 7 of 20 GLS did not show group differences.) Third, we expected higher values for those without than for those with dyslexia. This was based on the idea that especially the SLS can be assumed to have a factor structure. Those with dyslexia may be characterized by different combinations of symptoms as was shown in the previous study (Tamboer, Vorst, & Oort, 2014b), while those with non-dyslexia are characterized by none of the symptoms.

Table 7 provides all indices. Regarding the first expectation (higher values for Sample 2 than for Sample 1), we found higher values for the set of GLS in Sample 2 than in Sample 1, but the values for the set of SLS hardly differed between the two samples. This implies that knowledge about the purpose of the statements affects answer patterns on the GLS but not on the SLS. Regarding the second expectation (higher values for the SLS than for the GLS), we found higher values for the SLS than for the GLS in both samples, while this difference was very large in Sample 1. This implies that the SLS probably measure the same construct, while the GLS probably measure different constructs. Regarding the third expectation (higher values for non-dyslexic than for dyslexic students), we found that non-dyslexic students had higher values than dyslexic students in most cases. This implies that variability is higher in the dyslexic group compared with the non-dyslexic group, which was confirmed by the values of the mean of item variances. This might support the existence of a factor structure with those with dyslexia showing different combinations of response patterns on these factors.

All expectations were confirmed by the results. The SLS are less vulnerable than the GLS to answer patterns, while they probably better represent the same construct than the GLS. A factor structure may influence internal consistency reliability for the different dyslexic and non-dyslexic groups. We conclude that the SLS are more reliable predictors of dyslexia than the GLS, which confirms the conclusions regarding criterion validity. There are two remarks. First, internal consistency tends to be high in general when using many items as is the case here. Second, high values generally indicate that the same construct is measured by most items. However, the high indices of the SLS that were found here do not mean that the scale is unidimensional. The indices are a bit lower for dyslexic students than for non-dyslexic students, which results from higher item variances and points to the possibility of factor structures in the set of SLS.

Construct Validity

We hypothesized that dyslexia is most reliably predicted when it is taken into account that dyslexia is characterized by various symptoms. We therefore created a self-report inventory consisting of GLS and SLS. These SLS are based on a facet design with different facets representing different symptoms. We found that these SLS were better predictors of dyslexia than the GLS, that they were less vulnerable to answer bias and that analyses of internal consistency reliabilities supported the existence of a factor structure. Based on these findings, we further investigated construct validity of the self-report inventory by investigating whether the five-factor structure of the previous study (Tamboer, Vorst, & Oort, 2014b) could be confirmed. In this previous study, we found that five components with eigenwaardes above 1 explained 60% of variance in a large battery of tests and questionnaires. We named these components: spelling, phonology, short-term memory, rhyme/confusion and whole word processing/complexity.

In previous studies of self-report questionnaires (Bjornsdottir et al., 2013; Snowling et al., 2012; Wolff & Lundberg, 2014b), construct validity was investigated by performing exploratory factor analyses on all items that were available. In the present study, a principal components analysis of all 76 items resulted in 15 components in Sample 1 and in 14 components in Sample 2. A better solution was to perform exploratory factor analyses only on the items that were selected with various discriminant functions. When dyslexia and non-dyslexia were predicted with a small set of items, which are hypothesized to represent various symptoms of dyslexia, a factor structure should be visible even in these small sets of items. We only report the results of the functions that were acquired with the SLS and the combined set of GLS and SLS, because predictive validity and reliability of the GLS alone was relatively low. We performed exploratory factor analyses on the items that were selected for each discriminant function separately. In Table 8, all results are summarized. DF3, DF4 and DF5 were derived in Sample 1 and validated in Sample 2 so that analyses could be performed in both samples. All components have eigenwaardes above 1.

The most important finding was that one discriminant function with high criterion validity (DF4) showed the highest construct validity. This function (acquired on the basis of a test criterion) consisted of a set of 17 items (6 GLS + 11 SLS) and exhibited a five-factor structure explaining 60% of all variance (all components with eigenvalues above 1). This result confirms the result of our previous study (Tamboer, Vorst, & Oort, 2014b), in which we found a five-factor structure explaining 60% of variance as well. We studied the rotated component matrix of the five-factor structure found in the present study and decided that the following preliminary names for the five components were justified.

Component 1: whole word processing/complex words
Component 2: phonology
Component 3: confusion
Component 4: spelling/reading
Component 5: short-term memory/paying attention

There is some clear similarity between the two factor structures. Whole word processing, phonology, spelling/reading and short-term memory were found in both structures. It is however unclear how paying attention is related to short-term memory. Furthermore, rhyme and confusion are different constructs, although they may just as well represent the same construct. The main question now is why other discriminant functions resulted in less than five components. We further investigated the five-factor structure of DF4 by comparing this structure with the factor structures of other functions by correlating the factor scores of DF4 with factor scores of the other functions in Sample 1.

Correlations between the five-factor scores of DF4 and the four-factor scores of DF8 in Sample 1 could tell us something about how the use of a test criterion and a biographical criterion resulted in different factor structures. We found three high and significant (at 0.01 level) correlations: 0.74 between component 1 (whole word processing/complex words), 0.92 between component 2 (phonology) and 0.60 between component 4 (spelling/reading). Component 3 showed only a few low but significant (at 0.01 level) correlations, and component 5 of DF4 showed no correlations with components of DF8. This means that the fifth component (short-term memory/paying attention) is found when dyslexic and non-dyslexic students are selected with a test criterion but not when they are selected with a biographical criterion.

Correlations between the five-factor scores of DF4 and the two factor scores of DF3 in Sample 1 could tell us something about how the use of only SLS as predictors and the combined set of GLS and SLS as predictors resulted in different factor structures. We found two high and significant (at 0.01 level) correlations: 0.91 between component 1 (whole word processing/complex words) and 0.95 between component 2 (phonology). This means that using only the SLS as predictors resulted in a selection of items that represent whole word processing/complex words and phonology, but not confusion, spelling/reading and short-term memory/paying attention.

In Sample 2, DF4 resulted in a three-factor structure, explaining 54% of variance. An interpretation of the rotated component matrix learned that comparable items as in the other set of items are differently organized and more difficult to interpret. However, DF11 (acquired with a biographical criterion) resulted in a four-factor structure, explaining 57% of variance. An interpretation of the rotated component matrix here learned that interpretation was possible and that components of phonology, confusion, reading and short-term memory could be distinguished, while whole word processing/complex words were not clearly represented in this set of items.

Summarizing, we found support for a five-factor structure in a set of 17 predictors (6 GLS and 11 SLS) that were part of a discriminant function that was derived with a test criterion of dyslexia in Sample 1. With a biographical criterion, only four factors were found. Using only SLS as predictors resulted in only two factors. We conclude that different numbers of factors may originate from differences in answer patters, differences between GLS and SLS and differences between criteria used for the selection of dyslexic and non-dyslexic students. That the five-factor structure was found in Sample 1 makes sense because in this sample, it was unknown that dyslexia was assessed. This might have resulted in more variable answer patterns between items. That the five-factor structure was found with a test criterion of dyslexia makes sense because this criterion was considered to be more reliable, which might have contributed to the prevalence of those with dyslexia with more variable symptoms. This is consistent with the finding in the previous paragraph that the groups of non-dyslexics showed higher internal consistency values than the groups of dyslexics. This difference was most prominent for the comparison between those with and without dyslexia according to the test criterion, which implies that those with dyslexia are characterized by variable combinations of symptoms leading to variable answer patterns.

Convergent and Divergent Validity

Construct validity was further investigated with convergent validity and divergent validity. Of the six previous studies, only Snowling et al. (2012) thoroughly investigated convergent and divergent validity. As measurement of dyslexia, these researchers computed factor scale scores by summing the scores on the questions that had been identified as defining each factor. For comparisons, they furthermore used five raw scores on various measures. To be able to compare our results with those of the study of Snowling et al., we performed a similar analysis. Because test scores were only available in Sample 1, our analyses were restricted to this sample. As measurement of dyslexia, we only used the 56 SLS because the GLS showed low internal consistencies. We summed these items according to the original factors (spelling, phonology, short-term memory, rhyme/confusion and whole word processing/complexity) as were acquired in our previous study (Tamboer, Vorst, & Oort, 2014b). We performed correlational analyses between these scores and various test scores. See Table 9. Language-related tests are assumed to measure one or more symptoms of dyslexia. General cognitive tests are usually assumed to measure cognitive abilities that are not related to dyslexia.

All coefficients above 0.10 (and one of 0.09) are statistically significant because of the large sample size. However, Snowling et al. stated, '… it is the pattern of correlations which is important.' (p. 8).

Regarding convergent validity, Snowling et al. found high or moderate correlations between two scales (Reading Scale and Word Finding Scale) and three measures of language abilities (Spelling, Word Reading and Non-word Reading). We found mostly low and a few moderate correlations. The highest correlation was between the scale Spelling and measures of Dutch Dictation (0.43). A considerable amount of correlations above 0.20 was found between all factor scales and most language-related tests, except Sound Deletion.

Regarding divergent validity, Snowling et al. found low correlations between the two scales and a measure of block design. We used seven measures of cognitive abilities that are often regarded to be independent from dyslexia. Nevertheless, most scales correlated with Vocabulary, Verbal Analogies, Speed of Calculation and Hidden Figures, of which the last one is most surprising. However, the scales did not correlate at all (except for three coefficients of 0.10) with Conclusions, Numeric Progressions and Raven Progressive Matrices.

The overall pattern of the results is that low and a few moderate significant correlations were found between the scales of the inventory and measures that represent language abilities, while no correlations were found between the scales of the inventory and measures of general intelligence. Although there are a few unexpected coefficients (correlations between Speed of Calculation and all scales of the inventory; very low correlations between Sound Deletion and the scales of the inventory), the general conclusion is that both convergent and divergent validity are relatively high.

DISCUSSION

The aim of this study was to examine whether a self-report inventory of dyslexia is suited to use as a diagnostic instrument for identifying dyslexia and non-dyslexia. We found that both a test criterion and a biographical criterion can be used for reliable predictions with high criterion and construct validity. The best predictions are acquired when both general and specific language self-report statements are used as predictors. All together, not more than 20 items were needed. Our findings support findings from previous studies (e.g. Bjornsdottir et al., 2013; Snowling et al., 2012). Some alterations compared with these studies were found to be successful. However, some methodological issues are still not completely resolved yet. We will discuss this study from three perspectives: methodological, theoretical and practical.

Methodological: Predictive Validity

Using two different criteria of dyslexia and using less than 20 language statements as predictors, we found classification accuracies of more than 95%, although a validation of the best discriminant function from one sample to another sample resulted in lower classification accuracy of 85%. We suggested that sample differences may have reduced classification accuracy. In Sample 1, the GLS and the SLS were administered on separate days and without mentioning that dyslexia was assessed. In Sample 2, the GLS and the SLS were administered in one questionnaire, while the participants were aware that the statements were aiming to assess dyslexia. We investigated possible differences between samples in answer bias with analyses of internal consistency reliability. Indeed, we found that values of internal consistency reliability of the general statements were higher in Sample 2 compared with Sample 1. This may have resulted in lower classification accuracy when predictions of one sample are validated in a second sample.

Although we found high criterion validity, interpretation only makes sense when the nature of different criteria is accounted for. In each classification study of dyslexia, the question is what to trust more: a criterion or a prediction. When dyslexia is considered to be a disorder characterized by reading and spelling difficulties and when self-report statements only cover these difficulties, high criterion validity can be expected. In this study, we used a test criterion that was based on the view of dyslexia as a multi-dimensional cognitive deficit. We furthermore created self-report statements on the basis of a facet design that reflects many different symptoms. The most compelling improvement in our study was that especially statements about specific difficulties showed high predictive value, although a combination of general and specific statements resulted in the highest classification accuracies.

Nevertheless, also in this study, it cannot be determined with absolute certainty whether the test criterion or the best prediction represents the true dyslexic and non-dyslexic students. Comparing predictions with a test criterion, we found a negative predictive value of 99%, which means that almost all predicted non-dyslexic students are non-dyslexic according to a test battery as well. We found a positive predictive value of 89%, which means that 11% of the predicted dyslexic students are not dyslexic according to a test battery. However, the dyslexic students in our sample are intelligent, are highly educated and—in most cases—received remedial teaching at school. Some students may have developed coping strategies resulting in high performances on tests. Thus, it might just as well be that the self-report inventory is more reliable than a test battery.

Theoretical: Construct Validity

Construct validity of the self-report inventory appears to be good. The best classifying discriminant function consisted of 17 items, which exhibited a five-factor structure explaining 60% of all variance. We named the components as whole word processing/complex words, phonology, confusion, spelling/reading and short-term memory/paying attention. The reliability of this structure is supported by the fact that in a previous study (Tamboer, Vorst, & Oort, 2014b), a five-factor structure explaining 60% of variance was found as well. Construct validity was further supported by convergent and divergent validity. Sum scores that were based on the five-factor structure were correlated with measures of various tests that are known to be influenced by dyslexia and measures of various tests that are known to be unrelated to dyslexia. Although correlations supporting convergent validity were only moderate to low (but significant), the discrepancy with the very low correlations representing divergent validity was clear.

The five-factor structure was found using a test criterion. With a biographical criterion, a four-factor structure was found in both samples. Correlational analyses between the various factors of the different structures suggested that one component (short-term memory/paying attention) could not be identified with a biographical criterion. This may be explained by the fact that deficits of paying attention or short-term memory are not well recognized as symptoms of dyslexia by highly educated dyslexic students.

All together, we factor analysed the items of six different discriminant functions. Interpretations of these factors are difficult because some factors were only represented by a few items, while some items were related to more than one component of the analysis. Thus, we cannot say for sure whether the preliminary names are correctly representing the same construct. However, a theoretical conclusion here can be that both criterion validity and construct validity are good using both a test criterion and a biographical criterion and that any criterion should be based on more than spelling and reading difficulties alone.

Practical: Extern Validity

For purposes of diagnosing dyslexia, useful indices are the positive and negative predictive values. We found that negative predictive values were higher (97–99%) than positive predictive values (88–89%). This means that when a person is predicted as non-dyslexic, he or she is non-dyslexic with a very high probability. However, when a person is predicted as dyslexic, there is a chance of about 11–12% that this is not true. Thus, when the outcome of the self-report inventory is that a person is dyslexic, this person should be told that there is no absolute certainty and recommended to take steps for further investigation.

An important feature of this study is that the criteria did not exclude students based for instance on test performances. In other words, we did not distinguish between severely and moderately impaired dyslexic students. This supports the extern validity of the results. However, a shortcoming of this study is that we only used students who are highly intelligent, well educated and raised by parents with, on average, high socio-economical status. It thus remains unclear whether the same predictions can be applied on a more general population. On the other hand, using students may also have provided good insights in construct validity compared with when a general population would have been used. Students may have developed highly advanced coping strategies, which might remain undetected in a general population. We will further investigate this in a future study (in preparation). In this study, we hypothesize that in the case of dyslexia, self-report statements may be less vulnerable to effects of coping strategies or answer bias than tests. The idea is that any self-assessment of perceived difficulties may be carried out by comparing these difficulties to other cognitive skills and to the skills of friends and family with the same intelligence, schooling and socio-economical status. Self-report statements do not depend on intelligence, although the statements should be so that they are understandable for everybody.

CONCLUSION

We found support for the diagnostic value of self-report assessment of dyslexia. Highest classification accuracy was achieved with a criterion that was based on an extended test battery and that was consistent with biographical information. Construct validity was supported with factor analyses exhibiting a five-factor structure. For practical purposes, an important finding was that also only biographical information can provide a reliable and valid criterion of dyslexia. We found that self-report statements of dyslexia with the highest predictive value are characterized by specific descriptions of language difficulties, while more general statements suffer from answer bias. Extern validity will be investigated in a follow-up study.

APPENDIX A

Biographical questions

1. Are you dyslexic?
2. Did you, as a child, experience difficulties in reading?
3. Did you, as a child, experience difficulties in spelling?
4. Do you currently experience difficulties in spelling?
5. Do you currently experience difficulties in reading?
6. Do you currently experience difficulties in writing?

GENERAL LANGUAGE STATEMENTS

1. At school I preferred languages.
2. Every week I read a book.
3. As a child I did not like reading.
4. I am a quick reader.
5. When pronouncing difficult words, I stutter.
6. I often mix up expressions.
7. I find it easy to explain something to someone.
8. I experience difficulties recalling words.
9. I rarely experience difficulties in finding the correct spelling.
10. When I am writing, I often exchange letters.
11. I easily understand rules of grammar.
12. I find it difficult to write in an organised manner.
13. I like to play word games.
14. I can easily remember faces.
15. At school I always paid attention.
16. I find serial learning easy.
17. I easily learn a foreign language.
18. I experience difficulties in reading in English.
19. I do not like that English sounds differently than it is in written.
20. When I have to write in English, I experience difficulties in finding the right spelling of words.

Specific language statements

Reading

1. I sometimes skip a letter and read a different word.
2. I have no problems with reading difficult Dutch words.
3. Because I exchange some things during reading, I understand something differently.
4. When I read a difficult sounding word for the first time, I do not experience any difficulties.
5. Words which look alike when written, I rarely exchange during reading.
6. When a sentence has a subordinate clause, I can easily read it.
7. I read a different word that sounds almost the same as the written word, but that has a different meaning.
8. Sometimes when I read a text, it appears that I have read more than is written.

Writing

9. When I am writing, I know when to use a 'v' or an 'f'.
10. When I am writing something, I accidentally added superfluous words which made the text hard to understand for others.
11. When I am writing something, I rarely exchange letters (for instance 'eo' i.p.v. 'oe').
12. When I am writing something, I rarely forget to write certain words.
13. When I am writing down a word, I rarely think about its spelling although I know that I often write the word incorrectly.
14. Words such as 'universitair' and 'principieel' are not difficult for me to write.
15. When I am writing a sentence, I sometimes repeat parts of it.
16. When I am writing a sentence, words are in mixed order.

Speaking

17. When I pronounce a difficult sound for the first time, I make no mistakes.
18. During speaking, I exchange sounds such as 'psi' and 'spi'.
19. During a conversation, it is not hard for me to pronounce new words correctly.
20. During speaking, I have a correct choice of words.
21. During a conversation, I use complex sentences correctly.
22. When I speak, I exchange some parts of sentences.
23. During a conversation I sometimes struggle with the correct pronunciation of words.
24. During a conversation, I accidentally skip sentences which make it hard to understand what I mean to say.

Listening

25. I immediately hear it when somebody pronounces a sound incorrectly.
26. I hardly hear any difference between a 'p' and a 'b' in a word.
27. During a conversation, I sometimes notice that I perceived a sentence in the wrong order.

28. During listening, I hear the words exactly as they are used by the speaker.
29. When somebody reads out aloud something to me, I have a hard time understandings parts that resemble each other.
30. I understand a story exactly as somebody told it.
31. Even when I did not hear something during listening, I understand what somebody means.
32. When I am listening to a story, I accidentally add something for myself which makes the story difficult to understand.

Copying

33. During transcribing, it is rarely necessary to verify whether I write the correct letters.
34. When I am transcribing something, I rarely forget a letter in a word.
35. When I have to transcribe a word repeatedly, I sometimes write it down differently.
36. When I am transcribing something, I exchange letters with similar sounds (for example 'vrede' and 'wrede').
37. When I transcribing a sentence, I make no mistakes.
38. I transcribe something else than is written, which makes the text difficult to understand.
39. During transcribing, I sometimes write a different word that resembles the word which is written.
40. I can easily transcribe a difficult text.

Dictation

41. During a dictation, I rarely hesitate whether a word should be written with 'au' or with 'ou'.
42. When making a dictation, I almost automatically write down the words without mistakes.
43. Also when somebody uses extended words during a dictation, I immediately understand them while I am writing them down.
44. When I have to write down a word more than once during a dictation, I sometimes write them with different spelling.
45. When I want to write down literally what somebody says, I accidentally write down a different sentence because I think I heard it this way.
46. During a dictation, I easily write down long sentences.
47. When I make notes of a lesson, I forget some letters because I could not hear them.
48. A dictation usually goes by so quickly that I exchange parts of the text.

Reading aloud

49. When I have to read out aloud, I exchange words with similar pronunciation.
50. When I read out aloud something for the first time, I correctly pronounce difficult words.
51. When I read out aloud a text, I easily pronounce sentences in the correct order.
52. During reading out aloud, I skip parts so that others do not understand it anymore.
53. When I read aloud words with peculiar sounds, others can perfectly understand me.
54. During reading aloud, I rarely make up sentences which are not written down.
55. During reading aloud, I exchange words that look like each other.
56. When I read out aloud, I accidentally alter a sentence so that a new sentence arises.

Note that:

Litteral translations from Dutch to English are presented. The sentences represent items and examples of situations typical for the Dutch language and the Dutch aspects of dyslexia. Dyslexia research in other languages should incorporate aspects and situations that are typical for their own language.

APPENDIX B

Items of DF3 (12 SLS, selected with test criterion in Sample 1):
SLS: 2, 4, 12, 14, 17, 26, 35, 36, 42, 43, 44, 45.

Items of DF4 (6 GLS + 11 SLS, selected with test criterion in Sample 1):
GLS: 2, 10, 13, 15, 19, 20. SLS: 2, 12, 14, 17, 26, 35, 36, 42, 43, 44, 45.

Items of DF5 (4 BQ + 2 GLS + 9 SLS, selected with test criterion in Sample 1):
BQ: 1, 2, 3, 4. GLS: 10, 13. SLS: 10, 11, 14, 17, 35, 36, 38, 42, 45.

Items of DF7 (16 SLS, selected with biographical criterion in Sample 1):
SLS: 2, 4, 9, 12, 14, 16, 17, 18, 20, 28, 33, 35, 39, 42, 44, 45.

Items of DF8 (3 GLS + 15 SLS, selected with biographical criterion in Sample 1):
GLS: 10, 11, 14. SLS: 2, 4, 11, 14, 16, 17, 18, 20, 26, 35, 39, 41, 42, 44, 45.

Items of DF10 (10 SLS, selected with biographical criterion in Sample 2):
SLS: 1, 8, 11, 20, 30, 40, 42, 47, 54, 56.

Items of DF11 (5 GLS + 11 SLS, selected with biographical criterion in Sample 2):
GLS: 2, 3, 4, 14, 19.
SLS: 1, 11, 18, 20, 30, 35, 40, 42, 52, 54, 56.

REFERENCES

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.

Bjornsdottir, G., Halldorsson, J. G., Steinberg, S., Hansdottir I., Kristjansson, K., Stefansson, H., & Stefansson, K. (2013). The adult reading history questionnaire (ARHQ) in Icelandic: Psychometric properties and factor structure. *Journal of Learning Disabilities*, *20*, 1–11.

Bosse, M.-L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104*, 198–230.

De Groot, A. M. B., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, *33*, 600–629.

Furlanetto, L. M., Mendlowicz, M. V., & Bueno, J. R. (2005). The validity of the Beck Depression Inventory-Short Form as a screening and diagnostic instrument for moderate and severe depression in medical inpatients. *Journal of Affective Disorders*, *86*, 87–91.

Kramer, D., & Vorst, H. C. M. (2007). Communicatievragenlijst voor de opsporing van dyslectische kinderen (Communication Questionnaire for the Identification of Dyslexic Students). (Unpublished thesis). University of Amsterdam: Amsterdam, Netherlands.

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the Adult Reading History Questionnaire. *Journal of Learning Disabilities*, *33*, 286–296.

Mortimore, T., & Crozier, W. R. (2006). Dyslexia and difficulties with study skills in higher education. *Studies in Higher Education*, *31*, 235–251.

Ramus, F., & Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology*, *29*(1–2), 104–122.

Schulte-Körne, G., Deimel, W., & Remschmidt, H. (1997). Can self-report data on deficits in reading and spelling predict disability as defined by psychometric tests? *Reading and Writing: An Interdisciplinary Journal*, *9*, 55–63.

Soto, C. J., & John, O. P. (2009). Ten facet scales for the Big Five Inventory: Convergence with NEO PI-R facets, self-peer agreement, and discriminant validity. *Journal of Research in Personality*, *43*, 84–90.

Snowling, M., Nation, K., Moxham, P., Gallagher, A., & Frith, U. (1997). Phonological processing skills of dyslexic students in higher education: A preliminary report. *Journal of Research in Reading*, *20*, 31–41.

Snowling, M., Dawes, P., Nash, H., & Hulme, C. (2012). Validity of a protocol for adult self-report of dyslexia and related difficulties. *Dyslexia*, *18*, 1–15.

Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2014a). Identifying dyslexia in adults: An iterative method using the predictive value of item scores and self-report questions. *Annals of Dyslexia, Journal of Learning Disabilities*, doi:10.1177/0022219414558123.

Tamboer, P., Vorst, H. C. M., & Oort, F. J. (2014b). Five describing factors of dyslexia, *64*, 34–56.

Vorst, H. C. M. (2006). *Dyslexie Opsporingstest, DOT 7/8*. Handleiding voor leerkrachten. Intern Rapport. Amsterdam: Department of Psychology, University of Amsterdam.

Vorst, H. C. M., & Bermond, B. (2001). Validity and reliability of the Bermond–Vorst Alexithymia Questionnaire. *Personality and Individual Differences*, *30*, 413–434.

Wolff, U. & Lundberg, I. (2003). A technique for group screening of dyslexia among adults. *Annals of Dyslexia*, *53*, 324–339.

World Health Organization (2015). *The ICD-10 (Version: 2015) classification of mental and behavioural disorders: Diagnostic criteria for research*. World Health Organization: Geneva.