Screening for dyslexia in university students

Cavalli, Eddy; Brèthes, Hélène; Lefèvre, Elise; El-Ahmadi, Abdessadek; Duncan, Lynne G.; Bianco, Maryse

Published in:
Archives of Clinical Neuropsychology

DOI:
10.1093/arclin/acad103

Publication date:
2024

This is a pre-copyedited, author-produced version of an article accepted for publication in Archives of Clinical Neuropsychology following peer review. The version of record Eddy Cavalli, Hélène Brèthes, Elise Lefèvre, Abdessadek El Ahmadi, Lynne G Duncan, Maryse Bianco, Jean-Baptiste Melmi, Ambre Denis-Noël, Pascale Colé, Screening for Dyslexia in University Students: a Standardized Procedure Based on Conditional Inference Trees, Archives of Clinical Neuropsychology, 2024;, acad103, https://doi.org/10.1093/arclin/acad103 is available online at: https://academic.oup.com/acn/article-abstract/doi/10.1093/arclin/acad103/7517661.

Author Accepted Manuscript

Screening for dyslexia in university students: A standardized procedure based on conditional inference trees

Eddy Cavalli[1*], Hélène Brèthes[2], Elise Lefèvre[1], Abdessadek El Ahmadi[3], Lynne G. Duncan[4], Maryse Bianco[5], Jean-Baptiste Melmi[2], Ambre Denis-Noël[6], & Pascale Colé[2,7,8]

[1] Psychology and Cognitive Sciences, Laboratoire d'Etudedes Mécanismes Cognitifs, Université Lumière Lyon 2, Lyon, France

[2] Cognitive Psychology, Laboratoire de Psychologie Cognitive, Aix-Marseille Université & CNRS, Marseille, France

[3] Cognitive Neurosciences, Laboratoire de Neurosciences Cognitives, Aix-Marseille Université & CNRS, Marseille, France

[4] Psychology, School of Social Sciences, University of Dundee, Dundee, UK

[5] Educational Sciences, Laboratoire de recherche sur les Apprentissages en Contexte, Université Grenoble Alpes, Grenoble, France

[6] Psychology, Complexity and Cognition Laboratory, Université Côte-d'Azur, Nice, France

[7] Fédération de Recherche 3C (Cognition, Comportement et Cerveau), Aix-Marseille Université and CNRS, Marseille, France

[8] Institute of Language, Communication and the Brain, Aix-Marseille Université and CNRS, Aix-en-Provence, France

[*] Corresponding author at: Psychology and Cognitive Sciences, Laboratoire d'Etude des Mécanismes Cognitifs, Université Lumière Lyon 2, Lyon, France. E-mail address: eddy.cavalli@univ-lyon2.fr

Abstract

Objective: The focus of this study is on providing tools to enable researchers and practitioners to screen for dyslexia in adults entering university. The first aim is to validate and provide diagnostic properties for a set of seven tests including a one-minute word reading test, a two-minute pseudoword reading test, a phonemic awareness test, a spelling test, the Alouette reading fluency test, a connected-text reading fluency test, and the self-report adult reading history questionnaire (ARHQ). The second, more general, aim of this study was to devise a standardized and confirmatory procedure for dyslexia screening from a subset of the initial seven tests. We used conditional inference tree analysis, a supervised machine learning approach to identify the most relevant tests, cut-off scores and optimal order of test administration. **Method**: A combined sample of 60 university students with dyslexia (clinical validation group) and 65 university students without dyslexia (normative group) provided data to determine the diagnostic properties of these tests including sensitivity, specificity, and cut-off scores. **Results**: Results showed that combinations of 4 tests (ARHQ, text reading fluency, phonemic awareness, pseudoword reading) and their relative conditional cut-off scores optimize powerful discriminatory screening procedures for dyslexia, with an overall classification accuracy of approximately 90%. **Conclusions**: The novel use of the conditional inference tree methodology explored in the present study offered a way of moving towards a more efficient screening battery using only a subset of the seven tests examined. Both clinical and theoretical implications of these findings are discussed.

keywords: adults with dyslexia; diagnostic properties; supervised classification; conditional inference; validation procedure.

Screening for dyslexia in university students: A standardized procedure based on conditional inference trees

Introduction

Skilled reading has become an inescapable precondition for successful social and professional integration in highly literate societies. However, some individuals show significant learning difficulties, including developmental dyslexia. According to the *Diagnostic and Statistical Manual of Mental Disorders* 5th Ed. (DSM-5, American Psychiatric Association, 2013), developmental dyslexia (hereafter dyslexia) is a specific learning disorder characterized by inaccurate (or slow and effortful) decoding and word reading that may impair reading comprehension, as well as poor spelling skills, despite normal intelligence and appropriate educational opportunities.

For many individuals, dyslexia impedes their potential for academic and professional success, and some surveys including large cohorts of adults with dyslexia have shown that dyslexia in adulthood also has significant socioeconomic repercussions (MacDonald and Deacon, 2019). While a significant proportion of individuals with reading difficulties do not remain in education beyond secondary school (e.g., 8 out of 10 students according to a French survey conducted by de La Haye et al., 2008; see also Jonas, 2013) or are unemployed after leaving school (e.g., more than 14% of 16-24 year old in England; *Public Health England*, 2013), there is now evidence that an increasing number of individuals with dyslexia successfully manage to study at the university level despite persistent reading impairments (e.g., approximately 1.4% in France, *University Medical Service*; 4-6.3% in the UK, *UK Higher Education Statistic Agency*; 1.6-6.4% in Spain, López-Escribano et al. 2018; 1.5-4% in Sweden, Wolff and Lundberg, 2002).

Although more work is needed to address the numerous questions to be answered about dyslexia, school services have been implementing the identification and support of children and adolescents with dyslexia for many years (e.g., Shaywitz, Shaywitz, Fletcher & Escobar, 1990). The case of adults with dyslexia is rather different. In France, for example, while referral centers for learning disabilities are specifically dedicated to children and adolescents, there is no specific health service dedicated to adults with reading difficulties. Furthermore, most of our knowledge about dyslexia in adulthood comes from university students, who may constitute a very particular sample of individuals with dyslexia mainly because they have probably developed adaptive reading procedures to meet academic requirements (Colé, Cavalli, & Duncan, 2020). Nevertheless, epidemiological studies have shown that dyslexia is still a major barrier in the pursuit of higher education. For instance, in the UK, it has been reported in 2006 that students with dyslexia account for only approximately 3.2% of students entering university, although prevalence may be underestimated because the majority of dyslexic students (approximately 40–43%) remain undiagnosed before entering university (Warmington, Stothard, & Snowling, 2013). Moreover, students with dyslexia are more likely to withdraw from university studies and to attain poorer degree classes than their non-dyslexic peers (Bergey, Deacon, & Parrila, 2017; Richardson & Wydell, 2003). Therefore, there is a recognized need to identify university students with dyslexia and to develop diagnostic protocols that accurately assesses reading and reading-related skills to enable the implementation of intervention strategies that will ensure that university students with dyslexia achieve their academic potential.

Reading skills and reading-related skills in university students with dyslexia

Studies conducted with university students with dyslexia report persistent deficits in reading skills involving the identification of written words, decoding (reading pseudowords) and (text) reading fluency (see, for example, the meta-analysis by Swanson and Hsieh, 2009).

Such studies also show that diagnostic procedures for adults with dyslexia must differ in a number of respects from the methods used with children. While numerous symptoms of childhood dyslexia persist into adulthood, some characteristics of dyslexia change significantly with age. Adults develop compensations and adaptations induced by years of continuous exposure to written language, by a strong motivation to learn (Lefly & Pennington, 2000) and by both cognitive and socio-emotional protective factors (Haft, Meyers, & Hoeft, 2016). Complete compensation for reading difficulties across all components of reading is relatively rare (Erskine & Seymour, 2005; Parrila, Georgiou & Corkett, 2007) since basic processes in reading such as decoding seem to remain largely impaired in university student with dyslexia. Nevertheless, Parrila et al. (2007) further reported that 75% of their sample of 28 students with dyslexia showed reading comprehension performance in the typical range when speed was not taken into consideration (this figure dropped to 8% when speed was measured and taken into consideration in reading comprehension performance; see also Cavalli et al. (2019) and Deacon et al. (2012)). Indeed, reading comprehension tests are known to produce varying outcomes depending on the composition of the specific test, the prioritization of different (compensatory) reading strategies and the developmental level assessed (Cutting & Scarborough, 2006; Keenan, Betjemann & Olson, 2008; Deacon et al., 2012). Together these findings illustrate the need to move beyond reliance on norm-based tests of reading alone since factors such as compensation and the nature of underlying difficulties need to be taken into consideration for accurate and reliable diagnosis of dyslexia in adults.

While there is still an ongoing debate about the underlying cause(s) of developmental dyslexia (for a review, see Ramus and Ahissar, 2012; Peterson & Pennington, 2015), consensus has grown among researchers in recent decades that phonological processing is crucial for reading success, and a consistent relationship has been demonstrated between dyslexia and inefficient phonological processing skills (e.g., Vellutino, Fletcher, Snowling, & Scanlon, 2004; Lyon, 1995). It is well documented that individuals with dyslexia continue to show phonological impairments in adulthood, even relative to much younger reading age controls (Martin et al., 2010; Martin, Frauenfelder & Colé, 2013; Miller-Shaul, 2005). With regard to reading, phonological processing refers to a set of processes involving the phonological structure of words as used, for example, in pseudoword reading (which requires converting graphemes into phonemes), in phonological short-term memory (STM) (which retains the assembled phonological code of words or pseudowords to be decoded) and in Rapid Automatized Naming (RAN) (which requires the ability to rapidly access the phonological representations of spoken words). Phonological processing is also crucially involved in phonological awareness, which entails manipulating the phonological units (e.g., syllables, phonemes) of spoken words.

Although valuable, group level analyses alone cannot uncover the type of information that is critical for clinicians, namely, the proportion of adults with dyslexia who show a specific pattern of impaired performance on phonological processing measures. Using multiple single case study analysis, Ramus and colleagues (2003) reported that all adults with dyslexia in their study displayed impaired phonological processing skills on tests involving phonemic awareness, verbal short-term memory (VSTM) and rapid automatized naming (RAN) (see Cavalli et al., 2016, for comparable results). Similarly, in Parrila et al.'s (2007) study described above, impaired phonemic awareness was almost universally observed and Law, Vandermosten, Ghesquière and Wouters (2014) reported that 72% of adults with dyslexia exhibited a phonological awareness deficit, 53% a VSTM deficit and 31% a RAN deficit. These findings suggest that phonological deficits, already present in childhood (e.g., Snowling, 2000; Ramus et al., 2003), are persistent and prevalent in adulthood (e.g., Elbro, Nielsen, & Petersen, 1994; Martin, et al., 2010; Martin, et al., 2014). Finally, many studies report spelling deficits in adults with dyslexia using timed and untimed spelling tasks (for a review, see Tops and Brysbaert, 2020 and Swanson and Hsieh, 2009; and for a meta-analysis, see Reis et al., 2020). This is not surprising since spelling skills require the activation of knowledge about word pronunciation and the corresponding grapho-phonemic rules.

Screening for dyslexia in adults entering university

In order to determine whether or not an individual suffers from dyslexia through a diagnostic procedure, a distinction must be made between two types of investigation, namely screening and diagnosis confirmation. Screening should detect behavioral markers of dyslexia with a minimum of false negatives (i.e., high sensitivity), whereas diagnostic confirmation would focus on a minimum of false positives (i.e., high levels of specificity).

Diagnostic tools for adults have been developed in different languages. Several examples exist for English, namely, the *York Adult Assessment Battery Revised* (YAA-R; Warmington, Stothard, & Snowling, 2012), the *Bangor Dyslexia Test* (Reynolds & Caravolas, 2016), and the *Adult Screening Test* (DAST, Harrisson & Nichols, 2005). In other languages, there is the *Test for Automatized Spelling* in Dutch (TASP; Mostaert et al., 2016) and the *Test for Advanced Reading and Writing* (GL&SCHR; Depessemier & Andries, 2009), the computerized assessment battery called *BEDA* in Spanish (*Bateria de Evaluation de Dyslexia in Adultos*; Mejía, Diaz, Jimenez & Fabregat, 2012), the *One-Minute Reading Test for High-School Students* in Portuguese (Fernandes et al., 2017), the *SLS-Berlin* in German (Lüdtke, Froehlich, Jacobs & Hutzler, 2019), as well as Norwegian and Swedish batteries (Nergard-Nielsen & Eklund, 2017; Wolf & Lundberg, 2003). The diagnostic tools may differ from country to country but, as Obidzinski (2020) recently pointed out, the overall approach is comparable using measures of reading, spelling, reading-related skills and cognitive processes (e.g., working memory).

Less research has investigated tests that might be used for screening purposes to allow resource-limited health or disability professionals to focus more exhaustive diagnostic testing only on those who had already screened positively for dyslexia. Interestingly, previous studies in English conducted to identify screening tests for dyslexia in adults, suggest that only a small number of tests is required to reliably discriminate individuals with and without dyslexia (Hatcher et al., 2002; Swanson & Hsieh, 2009, Nicolson & Fawcett, 1997) rather than more time-consuming test batteries. Hatcher and colleagues (2002) developed a model, which categorized 96% of their small sample of students with dyslexia correctly based on the administration of only three screening tests from a much larger battery of 17 tests assessing aspects of literacy, processing skills, phonological skills and verbal fluency. The key tests were pseudoword passage reading (speed and accuracy), verbal short-term memory and writing speed. Spelling was also highlighted as an optional addition if further qualitative diagnostic information was required, although its inclusion would result in a higher rate of false positives. Similarly, Callens, Tops and Brysbaert (2012) conducted a large study comparing 100 Dutch-speaking university students with dyslexia and 100 control readers (i.e., skilled readers, matched on chronological age, sex and field of study). All students were assessed on a battery of 52 tests, consisting of an intelligence test, a memory test, and tests for vocabulary, processing speed, phonological awareness, rapid naming, arithmetic skills and reading (i.e., including word reading, pseudoword reading, text reading, text comprehension, and a word and sentence dictation task). In accordance with Swanson and Hsieh's (2009) meta-analysis on the characteristics of dyslexia in adulthood, Callens and colleagues found selective deficits in both speed and accuracy of performance in reading, writing, arithmetic, and phonological processing. Based on these results, Tops and colleagues (2012) used a cross-validation procedure to identify the minimum number of screening tests needed to validly and efficiently predict the presence of dyslexia in young adults entering higher education. The best predictive model they obtained using a stepwise logistic regression to fit a classification model yields a combination of three (out of 24) variables (i.e., predictors), including time-limited word reading accuracy, word spelling accuracy and a timed (reversal-matching) phonological awareness task, without loss of predictive power. This finding supports the conclusion that adding more tests did not improve the classification quality. Further, Tops et al. argued that this approach based on prediction is to be preferred over postdiction techniques due to its superior generalizability across samples.

The present study

This study has two main aims. The first is to validate a set of tests designed to discriminate individuals with and without dyslexia at university entry in France, as such measures are lacking. The second aim has a wider reach through the novel procedure that is used to determine the subset of these tests which forms the most efficient dyslexia screening procedure.

In France, there is a lack of validated tools and optimal diagnostic protocol for the classification of students with dyslexia (Cavalli et al., 2018). In this context, Cavalli and colleagues have recently validated the Alouette test for adults (Cavalli et al., 2018; see Rouweler et al. (2020) for a Dutch version of this test). This text reading test is considered as the "*gold standard*" test by both French-speaking practitioners and researchers in assessing reading level from childhood to adulthood due to the very good convergent validity and diagnostic properties of the test (Bertrand, Fluss, Billard & Ziegler, 2010; Cavalli et al., 2018; Sprenger-Charolles, 2019). Currently, the Alouette test is extensively used to identify dyslexia in French higher education as no other tests or diagnostic protocol are currently available with these diagnostic properties.

In line with previous work on French, Dutch and English (e.g., Callens et al., 2012; Cavalli et al., 2018; Hatcher et al., 2002; Swanson & Hsieh, 2009; Tops et al., 2012), and the findings of Parrila et al. (2007), seven tests were selected for investigation: the Alouette reading test (a text reading fluency test) (Lefavrais, 1967); the French translation of the Adult Reading History Questionnaire (i.e., French-ARHQ; Lefly & Pennington, 2000), as this test is widely used because of its excellent psychometric properties and ability to identify individuals with dyslexia (see Bjornsdottir et al., 2014; Deacon et al., 2012, 2017; Chevalier et al., 2017; Bergey et al., 2017); and five timed tests including word reading, pseudoword reading, connected text reading fluency, phonemic awareness, and spelling.

Our study will establish the detailed pattern of performance shown by a combined sample of university students with and without dyslexia on these seven tests in order to provide data from a normative group of readers (i.e., unimpaired readers) and a clinical validation group (i.e., impaired readers). These participants are readers of the French orthography, which has been shown to contain a higher percentage of complex multi-letter rules for reading than orthographies such as English and Dutch but which has a lower incidence of irregular words than either of these orthographies (Schmalz, Marinus, Coltheart & Castles, 2015). The objective is to determine the diagnostic properties of each test (sensitivity and specificity), along with the optimal cut-off value computed using the discriminatory outcomes provided by the receiver operating characteristic (ROC) curves method (for similar methodology, see Cavalli et al. 2018; see also Ter Huurne et al., 2023).

As already mentioned, from a clinical perspective and given the growing number of students with dyslexia entering higher education institutions in France and elsewhere, there is a need to develop reliable tools and standardized procedures to rapidly screen for individuals with reading difficulties. Accordingly, the second aim of this study is to explore which of these tests might be combined together to produce a standardized and confirmatory procedure to screen for dyslexia with good discriminatory power. In a complex learning disorder like dyslexia, which can manifest differently and is subject to compensation (Ramus et al., 2003; Colé et al., 2020), neither practitioners nor researchers are likely to be satisfied that a single test would constitute a very meaningful screening or diagnostic procedure. A combination of tests appears more powerful as a discriminatory procedure with a decision-making rationale based on the conjunction of two or more positive and negative responses (i.e., double-positivity and double-negativity, see Grenier, 1999). Additionally, the use of more than one test is likely to provide useful information to feed forward into future classification diagnosis, especially *explanatory* and *action-oriented* investigations designed to determine the appropriate support for individual students with dyslexia during their university studies (Tops & Brysbaert, 2020). As reviewed above, where authors have taken this approach by recommending shorter protocols, the results have provided a successful screening solution in a number of languages. While several studies have selected tests empirically from a larger pool on the basis of optimal discriminatory power, not all of these studies have explored a standardised procedure for administration (i.e., administration order, cut-off scores). Therefore, in this study we have attempted to provide empirically-based recommendations for dyslexia screening in the French language, which encompass a small set of validated tests together with a standardised procedure and precise cut-offs to apply in clinical assessment.

The novel methodology that we have adopted for this purpose has applicability beyond the French language and takes advantage of the popular conditional inference trees analysis, a supervised machine learning approach (Horthorn, Hornik, & Zeileis, 2006; 2015) to obtain supervised classification for and identification of the most relevant tests and cut-offs to screen for dyslexia using an optimal order of test administration. Conditional inference trees are one of the most widely used single-tree approaches, and provide unbiased variable selection thereby improving the reliability of variable selection (Hothorn, Hornik, Strobl, & Zeileis, 2015). To our knowledge, this decision-tree approach has not yet been used as far as classical tests to discriminate individuals with and without dyslexia in adults are concerned (but see Obidzinski, 2018, with memory tasks and Ritchie and Tuokko, 2011, in the dementia domain). Details about this procedure will be given in the *Data plan analysis* section.

Methods

Participants

One hundred and twenty-five university students participated in the study, including 60 students with dyslexia (66% female) and 65 students without dyslexia (63% female). While epidemiological studies typically demonstrate that dyslexia is more frequently diagnosed in males than females (see Arnett, Pennington, Peterson, Willcutt, DeFries & Olson, 2017), the university student population constitutes a special population in which sex differences may be reduced (this issue will be explored in more detail in the discussion to address the generalizability of our findings). In the current study, the female: male sex ratio of 1.9:1 is representative of the sex ratio favouring females in higher education (around 60%, i.e. 1.5:1 depending on subject area) and is also consistent with recent studies conducted on university students (e.g., 54%, i.e. 1.1:1 in Callens et al., 2012; 60%, i.e. 1.5:1 in Cavalli et al., 2018; 57%, i.e. 1.3:1 in the meta-analysis by Reis et al., 2020).

To be included in the present study, all participants with dyslexia (hereafter, referred to as dyslexic readers – DYS) and all participants without dyslexia (hereafter, referred to as skilled readers – SR) had to be university students, who were monolingual native French speakers, and who reported normal or corrected-to-normal hearing and vision. They were all recruited from the Universities of [*hidden for review*] and [*hidden for review*] to meet the requirements for representativeness of the target population. All the DYS participants were recruited and selected following an announcement sent to the Universities' medical services. All DYS had reported having experienced major difficulties in learning to read during childhood or adolescence and the persistence of these difficulties in adulthood had been confirmed by a physician in a specialized disability support service (*Mission Handicap)* of [*hidden for review*] Universities' medical services. Additionally, all DYS had received a formal diagnosis of dyslexia during primary school (mean age 9.1 years; SD=4.2) or in high school (up to age 17 years) according to the DSM-IV-TR criteria (American Psychiatric Association, 2000) and all had received remedial teaching during childhood (mean duration=4.5 years, SD=3.1). Each DYS participant was matched with a control participant on several variable including sex, chronological age (mean SR=20.4 years, SD=2.1; mean DYS=20.6 years, SD=2.3; t(123)=0.58; p=.56) and years of higher education (mean SR=1.1 years, SD=0.3; mean DYS=1.5 years, SD=0.9; t(123)=1.4; p=0.11). The two groups were also matched on academic program with approximately 60% of the participants enrolled in Social Sciences and Humanities and around 40% in Natural Sciences.

All participants were administered two neuropsychological tasks to estimate nonverbal IQ (Raven's matrices; Raven, Court & Raven, 1995) and verbal IQ (Echelle de Vocabulaire en Images Peabody, EVIP; Dunn, Thieriault-Whalen, & Dunn, 1993; the French adaptation of the Peabody Picture Vocabulary Test-Revised, PPVT-R; Dunn & Dunn, 1981; see Cavalli et al., 2016). All participants performed above the fifth percentile on both nonverbal IQ (mean SR=44.0, SD=6.9; mean DYS=43.4, SD=7.6; t(123)=0.53; p=0.59) and verbal IQ (mean SR=39.8, SD=4.3; mean DYS=38.6, SD=6.1; t(123)=0.58; p=.56), thereby confirming that none of the participants presented a deficit in nonverbal reasoning or in semantic oral language skills. Moreover, potential participants with self-reported or formal diagnosis of specific language impairment or other impairments that could impact language ability (e.g., autism spectrum disorder) were not included in the present study. In addition, none of the participants had any known neurological or psychiatric disorders, and all SR participants who reported having any known learning disorder or experienced difficulties in learning to read during childhood or adolescence were not included in the present study. Finally, five DYS participants were excluded from this study due to missing data or drop out.

The experiment was conducted in accordance with the Declaration of Helsinki and with the understanding and written consent of all participants. The study was approved by the [*hidden for review*] University ethics committee.

Tests

The following tests were administered to all participants: French version of the Adult Reading History Questionnaire (ARHQ, a self-report questionnaire; Lefly and Pennington, 2000); Alouette test (Lefavrais, 1967) a standardized French reading test for adults (Cavalli et al., 2018); as well as five tests we developed ourselves, namely, the one-minute word reading, two-minute pseudoword reading, phonemic awareness, spelling and text reading fluency tests. These latter five tests along with the French translation of the ARHQ are freely available and the full program (both the spelling and phonological awareness tests are computerized), instruction, items, and scoring for the French language can be found on the following *Open Science Framework* website https://osf.io/zmf82/. The remaining tests are commercially available. The time to complete all the tests (excluding both verbal and non-verbal IQ tasks) is estimated to be about 30 minutes. Moreover, training for screeners is expected to be very short.

ARHQ. The ARHQ is a 23-item self-report questionnaire developed by Lefly and Pennington (2000). This questionnaire has been translated for French (Brèthes et al., 2022). This scale is a useful instrument in clinical practice for screening for dyslexia in adults. It includes items on reading habits, reading and spelling abilities, reading speed, reading habits, attitudes toward school and reading, additional assistance received, repeating grades or courses and effort required to succeed, separately for elementary school, secondary school, post-secondary education, and current life (Deacon et al., 2012). A global score was calculated by totalling the points on the items and then dividing by the maximum possible score. Scores could range from a low of 0 to a high of 1 with the lower score indicating less reading difficulty. Reliability (Cronbach's α) was .87 (95% Confidence Interval [.83 - .89]). For comparison, reliability was .93 in Deacon et al. (2012), and .90 in the original version of Lefly and Pennington (2000). Moreover, Lefly and Pennington (2000) used a discriminant function analysis on the ARHQ scores and revealed a cut-off score of .40 (81.8% sensitivity, 77.5% specificity), and in the Icelandic version of the ARHQ (Bjornsdottir et al., 2014), a ROC curve analysis revealed a cut-off score of .43 (84.5% sensitivity, 83.7% specificity) to identify individuals with a history of reading difficulties indicative of dyslexia.

Alouette test (reading text fluency test). The Alouette test (Lefavrais, 1967) requires participants to read a 265-word text aloud as rapidly and as accurately as possible, with a 3-minutes time limit. The specificity of this test is that the text consists of real words in meaningless but grammatically and syntactically correct sentences, thus preventing dyslexic readers and poor readers from compensating for their written word recognition difficulties by using contextual information. The test yields measures of accuracy (A, number of words correctly read), reading time (RT, time taken to read the text), and reading efficiency (called CTL, computed by the following formula: CTL = (A/RT)*180, where A = accuracy (self-corrections included), and RT = reading time (maximum = 180 sec); see Bruyer and Brysbaert (2011), Cavalli et al. (2018), for a detailed presentation of efficiency scores).

One minute Word Reading. Participants were instructed to read aloud written words as fast and accurately as possible within one minute. Words were presented on a printed sheet containing 6 words by line 120 disyllabic French words with a length between 4 and 9 letters (mean=6.4; SD=1.29) and a frequency varying from low to high (mean=28.6; SD=43.4) were selected using the *lexique.org* database (New et al., 2001). 80% of the words were concrete and 20% were abstract (Bonin et al., 2018). An efficiency score was then computed for each participant by taking in account both accuracy (A) and reading time (RT): (A/RT)*60 (see Brèthes et al., 2022).

Two minute Pseudoword Reading. Participants were instructed to read aloud written 116 pseudowords as fast and accurately as possible within 2 minutes. Pseudowords were presented on a printed sheet containing 6 pseudowords by line. They were between one and two syllables long and had an average length of 5.5 (SD=0.5). Again, efficiency scores were calculated for each participant: (A/RT)*120 (see Brèthes et al., 2022).

Phonemic Awareness. In this computerized test, participants were instructed to pronounce, as fast and accurately as possible, the pseudowords they heard by deleting the first phoneme (e.g., they heard /blɔ/ and have to say /lɔ/). 30 monosyllabic pseudowords with a Consonant Consonant Vowel (CCV) structure were selected. Reliability (Cronbach's α) was .88 (95% Confidence Interval [.83 - .90]). As in the previous tasks, the final scores were efficiency scores taking into account both accuracy and response times: (A/RT)*100.

Spelling. In this computerized timed-test, participants were instructed to write down on a sheet the words they heard as accurately as possible (see Tops et al., 2012 for more details on the method). They had 3 seconds to write down a given word before hearing the next one and were instructed to go on if they could not write the word. They were also warned that they could not go back on a word to correct its spelling. Eighty words were selected from the *lexique.org* database (New et al., 2001). Words varied in spelling consistency (half consistent words, half inconsistent words), written frequency (mean=47; SD=100), were from 3 to 8 letters long (mean=6; SD=1,35) and were composed of 1 to 2 syllables. Words were previously recorded in a soundproof room by a French native speaker. Word order was randomized. The final score corresponds to the number of correctly written words (maximum 80).

Connected-text Reading Fluency. In this reading test, participants were instructed to read aloud a text selected from *"L'Écharpe de Soie Rouge"* (Leblanc, 1913) as fast and accurately as possible in 1 minute and in accordance with punctuation marks. It is a short narrative literary French text of 434 words and 24 sentences (reduced to 337 words and 17 sentences). The final score corresponds to the number of correctly read words in 1 minute.

Data Plan Analysis

For each of the administered tasks, DYS and SR groups were compared using *t* tests and Cohen's *d* effect sizes were computed to estimate the strength of the differences (Cohen, 1988).

ROC curves analyses. To determine the psychometric discriminative properties (sensitivity and specificity) and optimal cut-off value (which takes into account dyslexia prevalence) of all the tests administered, results from DYS and SR groups were examined using ROC curves analysis, i.e., an application of signal detection theory that provides a graphical representation of the performance of a binary classifier as discriminant thresholds vary (Swets, 1988). As an index of discriminant ability, we used the area under the curve (i.e., AUC index) which is an effective and combined measure of sensitivity and specificity that describes the inherent validity of a test, and therefore provides a metric of the utility of the instrument with AUC values of 1.0 indicating perfect discrimination and 0.5 indicating poor discrimination (see Cavalli et al. 2018, for similar methodology). Specifically, ROC curves plot sensitivity versus 1 – specificity to create a curve comparing the discriminative value of an instrument relative to no discrimination, informing on how the proportions of true positives and false positives change for each possible cut-off value. In the present study, AUC, specificity (Sp), sensitivity (Se), and dyslexia prevalence (pv; fixed at 5% as described in the literature; Swanson, 2012) were used to determine optimal cut-off values for all the assessment tests with their associated Positive Predictive Value (PPV, i.e., the probability that subjects with a positive screening test truly have the outcome of interest; computed with the formula PPV = f(Se, Sp, pv) and Negative Predictive Value (NPV, i.e., the probability that subjects with a negative screening test truly don't have the outcome of interest; computed with the formula NPV = g(Se, Sp, pv) (see Huang, Fong, Wei, and Feng, 2011, as concerned with estimates positive and negative predictive values).

Conditional inference trees analysis. To identify the most relevant tests and their respective thresholds in a diagnostic procedure, we then used the supervised machine learning classification method. The classification algorithm chosen was a conditional inference tree from the R package party (Hothorn, Hornik, & Zeileis, 2006). This method uses recursive partitioning of dependent variables based on the value of correlations. Recursive partitioning methods have become popular and widely used tools for nonparametric regression and classification in many scientific fields (for a comprehensive introduction on the principles of the standard recursive partitioning methods and recent methodological improvements, see Strobl et al. 2009). These methods allow the modelling and interpretation of the relationship between a categorical variable and various predictors. Their use addresses a problem often encountered in psychology studies where the number of participants is small compared to the number of predictors, which limits the use of prediction with linear models such as logistic regression. In that context, nonparametric classification methods such as conditional inference decision trees represent an attractive alternative as they can deal with large numbers of predictors variables even in the presence of complex interactions. Interestingly, decision trees constitute a framework for the creation of an explicit hierarchy of tests that results in a partitioning of the decision space and are therefore better suited for diagnostic purposes than the standard classification and regression trees. To help our statistical understanding, the principle of conditional inference trees can be summarized in the following 3 steps: Step 1 consists of selecting the predictor which helps best to distinguish between different values of the response variable, using some statistical criterion. In Step 2, a split is made in this variable, splitting the data into several data sets using binary partitioning algorithms. Lastly, Step 3 consists of repeating Steps 1 and 2 recursively until no further splits can be made, based on certain pre-defined criteria (for a more in-depth presentation and illustration of conditional inference trees methods, see Levshina, 2015).

Conditional inference decision trees have very good comprehensibility and form a structure made by nodes and branches, starting at a single root node and ending in the terminal nodes, also called the leaves of the tree (Sardá-Espinosa et al., 2017). Conditional inference tree algorithms perform predictor selection through statistical test permutation on the performance of data partition by each possible predictor and then choosing the best classification variable and its respective cut-off score[1], retaining it only if the predictor is adding significant classification power (Sardá-Espinosa, Subbiah, & Bartz-Beielstein, 2017). This selection, motivated by statistical tests makes it possible to stop the construction of the decision tree before reaching over-fitting, a serious issue with decision trees (Hothorn et al., 2006; 2015). In the present study, the cutoff was kept at $\alpha = 0.05$ with a Bonferroni correction. Moreover, the use of this classification algorithm has been paired with a ten-fold cross-validation (implemented in the *caret R* package; Kuhn et al., 2020; see Jung et Hu, 2015) to compute

[1] It must be noted that cut-off scores calculated within decision trees are only valid within the model and not as a replacement for cut-off scores calculated with the ROC method, since in this classification procedure, cut-off scores are conditional on the previous node.

both NPV and PPV indices and to assess the ability of the model to generalize to new data and therefore not overrate the performance of the trained algorithm.

Results

Statistical description

Table 1 provides performances and difference effect sizes (Cohen's *d*; Cohen, 1988) obtained by both DYS and SR groups on all the seven tests administered. In all the tasks, DYS' performances were significantly less efficient than that of the SR: all $t(123) > 5$ and $p < .001$. Moreover, effect sizes as computed with Cohen's *d* were large (all $d > .70$).

Please insert Table 1 about here

ROC curves analyses

Figure 1 presents the ROC curves with confidence interval (left part) and optimal criterion cut-off plots (right part) for efficiency scores of word reading, pseudoword reading, phonemic awareness, and spelling (Figure 1.A); and scores of text reading fluency, Alouette reading and ARHQ (Figure 1.B). The best balance between the proportion of true positives and true negatives correctly identified using the cut-off (i.e., sensitivity and specificity, respectively) was obtained by classifying individuals as dyslexic with the following cut-off scores (see Table 2): word reading score ≤ 99 (90% sensitivity, 69.2% specificity), pseudoword reading score ≤ 105 (93.3% sensitivity, 93.8% specificity), phonemic awareness score ≤ 1.6 (88.3% sensitivity, 81.5% specificity), spelling score ≤ 73 (78.3% sensitivity, 73.8% specificity), text reading fluency score ≤ 186 (96.7% sensitivity, 67.7% specificity), Alouette reading score ≤

426.6 (91.6% sensitivity, 92.3% specificity) and ARHQ score > 0.47 (93.3% sensitivity, 93.8% specificity).

Please insert Figure 1.A about here

Please insert Figure 1.B about here

Table 2 presents the classification data of the ROC curves with 95% confidence interval (CI) for all the assessment tests, along with the AUC index (with 95% CI and p value), with the associated sensitivity and specificity scores (and 95% CI).

Please insert Table 2 about here

For the seven assessment tests, the ROC curves analysis provided several pieces of information.

First, the AUC for word reading efficiency, pseudoword reading efficiency, phonemic awareness efficiency, spelling efficiency, connected-text reading fluency, Alouette reading efficiency, and ARHQ indicated high classification accuracy (all ps < .001).

Second, since the Alouette and phonemic awareness tests are the only two tests where measures of accuracy and response time are collected separately, we conducted pairwise comparison of ROC curves in order to determine which of the three measures (accuracy, speed or efficiency) represents the best discriminatory measure. For Alouette reading, the results showed that the AUC for the reading efficiency score was greater than the AUC for speed (difference between areas = .07 ± .03; p = .02) and accuracy scores (difference between areas = .08 ± .03; p = .01), therefore confirming previous results by Cavalli et al. (2018). Similarly, for phonemic awareness, the results showed that the AUC for the efficiency score was greater than the AUC for speed (difference between areas = .28 ± .04; $p < .001$) and accuracy scores (difference between areas = .30 ± .05; $p < .001$).

Third, based on their psychometric discriminative properties, three tests show excellent positive and negative predictive values (i.e., the pseudoword reading test with PPV = 93.3 and NPV = 93.7; the Alouette reading test with PPV = 91.6 and NPV = 93.7; and the ARHQ self-report questionnaire with PPV = 93.3 and NPV = 93.7), three tests present good results for these values (i.e., the word reading test with PPV = 87.3 and NPV = 74.5; the phonemic awareness test with PPV = 87.4 and NPV = 82.6; and the connected-text reading fluency test with PPV = 95.3 and NPV = 74.9), and one test presents significant but moderate or poor predictive values (i.e., the spelling test with PPV = 77.2 and NPV = 74.9).

Conditional inference tree analyses

The second objective of this study was to devise a standardized and confirmatory procedure to screen for dyslexia by the convergent use of a small number of tests based on their discriminative properties, using conditional inference trees analysis (Horthorn, Hornik, & Zeileis, 2006). Two classification models were run: in the first one we included the scores of the seven tests we administered; in the second one we excluded the ARHQ self-report questionnaire because this questionnaire (French translation of the ARHQ) and the associated normative data are still not available for French practitioners (i.e., psychologists and speech therapists). Moreover, while being widely used in other countries to identify individuals with a history of reading difficulties indicative of dyslexia, in this questionnaire the reading difficulties measured are self-estimated, and some practitioners may feel uncomfortable in using it.

The first decision tree presented in Figure 2, was built from 7 predictors (i.e., word reading, pseudoword reading, phonemic awareness, spelling, text reading fluency, Alouette reading, and ARHQ) and only 3 predictors were retained to be relevant (i.e., ARHQ, phonemic awareness and pseudoword reading). The overall accuracy of the classification was 92%, which is significantly better than the no-information rate ($p<.001$). The sensitivity of the model (i.e., the ability to correctly detect individuals with dyslexia) was 87% and the specificity (i.e., the ability to correctly detect skilled readers) was 92%. Moreover, the predictive values were excellent with PPV = 90.32 and NPV = 93.65. On the classification tree, we can see that the first node separates individuals according to their ARHQ scores with individuals who obtained a score of 0.47 or above most likely to be DYS while those who obtained a score inferior to this threshold were most likely to be SR. Further distinction between SR and DYS is achieved through phonemic awareness (node 2) and pseudoword reading (node 5) tests. On the one hand, performances on the phonemic awareness test (left part) allow a finer split among individuals who were most likely to be SR according to the ARHQ (ARHQ score of 0.47 or below). 98% of individuals who obtained phonological awareness efficiency scores above 1.06 are categorized as SR and 2% as DYS (node 4). For individuals who obtained phonological awareness efficiency scores of 1.06 or less, 57% are categorized as SR and 43% as DYS (node 3). On the other hand, performances on the pseudoword reading test (right part) allow a finer split among individuals who were more likely to be DYS according to the ARHQ (ARHQ score above 0.47). 100% of individuals who obtained a score of 105 or less are categorized as DYS and 0% as SR (node 6). For individuals who obtained a score above 105, 57% are categorized as SR and 43% as DYS (node 7). In summary, the classification tree correctly identifies DYS by following the double positivity path (nodes 1 – 5 – 6), and SR by following the double negativity path (nodes 1 – 2 – 4).

Please insert Figure 2 about here

Figure 3 provides information relative to the second classification model. The complete dataset contained 6 predictors (i.e., word reading, pseudoword reading, phonemic awareness, spelling, text reading fluency, and Alouette reading) and only 2 predictors were retained as relevant (i.e., Alouette reading and phonemic awareness). The accuracy of the built tree was 89.6%, which is significantly superior to the no information rate ($p<.001$). Similarly to the first model, the sensitivity was 86.67% and the specificity was 92.31%. However, the predictive values were excellent with a similar PPV = 91.23 but a different NPV = 88.24. The first node separates individuals according to their Alouette reading scores with individuals who obtained a score of 419.45 or less (left part) most likely to be DYS (node 2) while those who obtained a score superior to this threshold (right part) were most likely to be SR (node 5). Further distinction between groups is achieved through the phonemic awareness test, with two conditional cut-off scores. On the one hand, among individuals who had an Alouette score of 419.45 or less, 98% of those who obtained an efficiency score of 1.75 or less on the phonemic awareness test are categorized as DYS and 2% as SR (node 3). For individuals who obtained an efficiency score above 1.75, 57% are categorized as DYS and 43% as SR (node 4). On the other hand, among individuals who had an Alouette score above 419.45, 100% of those who obtained an efficiency score above 1.85 on the phonemic awareness test are categorized as SR and 0% as DYS (node 7). For individuals who obtained an efficiency score of 1.85 or less, 74% are categorized as SR and 26% as DYS (node 6). In summary, the classification tree correctly identifies DYS by following the double positivity path (nodes 1 – 2 – 3), and SR by following the double negativity path (nodes 1 – 5 – 7).

Please insert Figure 3 about here

Discussion

This study has two main goals. The first one is to validate a set of seven literacy-related tests according to the psychometric discriminative properties of each test for French-speaking University students. The second goal is more general because it proposes a methodology for determining an optimal and short protocol to screen for dyslexia using the conditional inference tree procedure. The outcome is designed to provide French-speaking researchers and practitioners with a set of tools for assessing students with dyslexia entering higher education as well as some initial proposals about how subsets of these validated tests might most effectively be combined to form a standardised screening procedure.

Identifying students with dyslexia in higher education

Both the dyslexic and the skilled reader group were assessed on each of the seven tests examined in the current study. Their respective mean performances were compared using group comparison analyses and Cohen's d effect sizes to estimate the strength of the differences, revealing moderate to strong effect sizes (i.e., all Cohen's d values were within the interval of [0.7 – 2.9]). Afterwards, the psychometric discriminative properties of each test (including cut-off values, sensitivity and specificity) were computed using ROC curves analysis which allowed us to conclude that three tests possess excellent discriminative properties with high Se and Sp values (i.e., the two-minute pseudoword reading test, the Alouette reading test, and the ARHQ self-report questionnaire), three tests had good discriminative properties for screening with high Se value (i.e., the one-minute word reading test, the phonemic awareness test and the connected-text reading fluency test) and one test presents moderate or poor but nevertheless significant discriminative properties (i.e., the spelling test).

It is not altogether surprising that the individual tests administered can reliably discriminate French-speaking adults with dyslexia from skilled readers at university. The outcome is in line with previous results reported by Brèthes et al. (2022), Cavalli et al. (2016; 2017; 2018), and Martin et al. (2010; 2014), showing that word and pseudoword reading, text reading fluency (the Alouette test), as well as phonemic awareness are still impaired among French-speaking university students with dyslexia (see also the meta-analysis by Swanson and Hsieh, 2009, for findings from other languages). The results obtained with the ARHQ self-report questionnaire further extend those of Cavalli et al. (2016) and Marchetti et al. (2023) by demonstrating that this questionnaire is very effective in identifying French university students with dyslexia. While there are no previous data on connected-text reading fluency and spelling in French for comparison, the connected-text reading fluency test results are consistent with numerous studies from other languages reporting impaired performance in university students with dyslexia (e.g., Covignou et al., 2019; Nergård-Nilssen & Hulme, 2014; Ré, Tressoldi, Cornoldi, & Lucangeli, 2011; Reis et al., 2020; Warmington et al., 2013). Only the outcome of the spelling task is unexpected, where the Cohen's d value showed that the dyslexics' spelling performance was only moderately impaired despite the consistent evidence across different orthographic systems of severely impaired spelling skills in adults with dyslexia (see below for further discussion of this point).

To our knowledge, few studies have used methods comparable to those of the present study by assessing the discriminatory power of *each* test to be validated (i.e., ROC curves analysis, sensitivity and specificity values, PPV and NPV values) or through the use of similar tests or materials. This is mainly because most of work reviewed above was concerned not with the psychometric discriminative properties of each of the tests contained within the assessment battery but rather with the properties of the battery as a whole (for example, Harrison and Nichols, 2005, Reynolds and Caravolas, 2016; Warmington et al., 2013).

However, there are a small number of comparable results for the Alouette test and for the word and pseudo-word reading tests. The Alouette test is a text reading fluency test, which requires the reading of meaningless but grammatically correct sentences and which until recently had no equivalent outside of the francophone countries where it is widely used by researchers and practitioners (e.g., France, Belgium, Switzerland). As it is considered a gold standard test (Bertrand, Fluss, Billard, & Ziegler, 2010), further research can shed important light on its value and application. Text reading fluency provides a more natural method of investigating reading skills than single word reading because words are almost never read in isolation. Nevertheless, by providing contextual cues, normal text may allow readers with dyslexia to compensate for their impaired decoding skills. Rouweler et al. (2020) recently argued that because the meaningless text of the Alouette test does not support the use of contextual cues, it offers an interesting way of assessing reading skills in a more ecological manner. Rouweler and colleagues found excellent sensitivity and specificity properties for a Dutch version of this test inspired by the Alouette text, the *Flamingo* test, in a university student population. Further, strong correlations were also evident with word and pseudoword reading tests (which independently had very good psychometric properties), thus attesting to convergent validity between the *Flamingo* test and reading skills assessed in more traditional ways and making the *Flamingo* test a valuable new diagnostic instrument to assess reading skills.

We were also interested to offer a test assessing connected-text reading fluency, which is defined as the ability to read a meaningful connected text, quickly, accurately and with appropriate intonation (Rose & Rouhani, 2012). Crucially, grouping words into syntactic units of meaning, making quick use of punctuation and choosing pauses and intonation to make a text meaningful are all necessary skills in such a task. Thus, connected-text fluency is considered as an index of reading ability (Fuchs, Fuchs, Hosp, & Jenkins, 2001), and a lack of fluency is one of the most salient characteristics of dyslexia (see Rose and Rouhani (2012) and Niergard-Nillsen and Hulme (2014) for findings from adolescents and adults with dyslexia, respectively). Our data show that a connected-text reading fluency test provides a reliable indicator of reader category among adults in the first set of results to be reported on French-speaking adults with and without dyslexia. Although some previous studies have examined connected-text reading fluency, they did not provide psychometric properties for this test (Gola-Asmussen, Lequette, Pouget, Rouyer & Zorman, 2011; Rose & Rouhani, 2012).

With regard to the French translation of the ARHQ, our results are consistent with those reported by Deacon et al. (2012) in English, Bjornsdottir et al. (2014) in Icelandic, Wolff and Lundberg (2003) in Swedish, Alves and Castro (2005) in Portuguese, and Maurer (2005) in German; all of whom showed that the administration of a self-reported questionnaire of reading difficulties can also be considered as a reliable indicator of dyslexia.

Our spelling test showed only moderate discriminative properties which contrasts with the numerous studies reporting strongly impaired spelling skills in university students with dyslexia (in Spanish, López-Escribano et al., 2018; and in English, Peter et al., 2020, and Warmington et al., 2013) and in French-speaking children with dyslexia (Gosse & Van Reybroeck, 2020, Blampain et al., 2021). Indeed, Nergård-Nilssen and Hulme (2014) concluded that as spelling impairments are the most prominent residual behavioural marker of dyslexia in adults, a spelling test should be a stable component of dyslexia diagnostic procedure. The moderate discriminative properties of the spelling test in our study suggest that this test should not systematically be part of a diagnostic protocol for the classification of French students with dyslexia, although data for French-speaking adults are clearly lacking (but see Chetail et al., 2019). This outcome is puzzling since the phoneme-grapheme correspondences central to spelling tasks are rather inconsistent in French (Peereman, Lété and Sprenger-Charolles, 2007; Peereman, Sprenger-Charolles & Messaoud-Galusi, 2013), leading to the expectation that spelling should be impaired in French-speaking dyslexics. Furthermore, a meta-analysis by Reis et al. (2020), which took account of the impact of the opacity of the orthographic system, confirmed the persistence of spelling deficits among adults with dyslexia and also showed that intermediate and opaque orthographies lead to stronger effect sizes relative to transparent orthographies (but see Georgiou et al. (2021) for a meta-analysis of orthographic skills more generally, which corroborates the presence of adult deficits but not the influence of orthographic consistency). The spelling test we chose may not have been sensitive enough to show a clear effect of dyslexia impairment. Indeed, Ré et al. (2011) observed that a spelling test identical to ours only differentiated Italian adult readers with dyslexia from skilled readers if it is administered in conjunction with articulatory suppression (i.e., by asking them to repeat aloud the syllable 'la' continuously during the dictation), which interferes with (phonological) processing capacity. However, while the task was designed to tap speed of processing by presenting the dictated word in a limited time (i.e., constant pace of one word per 3 seconds in a pre-recorded audio file), one cannot exclude the hypothesis of a speed-accuracy trade off among the control group who showed very high accuracy scores (i.e., close to ceiling effect) reducing the variance in spite of the timed component.

Thus, on the basis of our first set of analyses of the discriminatory power of each test, we would propose that the best combination of tests to be used in a short protocol to identify the presence of dyslexia in French university students would be the two-minute pseudoword reading, the Alouette test and the ARHQ, although the one-minute word reading and the phonemic awareness tests that we developed also have good discriminatory power. This differs from the composition of other short protocols for dyslexia screening developed for English using pseudoword passage reading, verbal short-term memory and writing speed (Hatcher et al., 2002) and for Dutch comprising word-reading, spelling and phonological awareness skills (Tops et al., 2012), probably reflecting differences in the languages such as orthographic depth and/or in the methodologies used to design these tools. For example, the study by Tops et al. (2012) did not assess the discriminatory power of the three individual tests focusing instead on the discriminatory power of the protocol as a whole.

The psychometric discriminative properties of a test may not always be the best indicators of whether this test should be included in a screening protocol. Tests may present excellent discriminative properties individually but when combined may prove redundant due to intercorrelations with other tests or conversely, tests may add independent discriminatory power to a screening procedure despite possessing only moderate discriminative properties individually. The next section details the outcome of our use of a novel methodology to identify the optimal combination of our validated tests in designing a short but reliable screening tool, which specifies the properties of the tool as a whole together with a standardized procedure and appropriate cut-offs for administration.

Methodology for the design of a dyslexia screening procedure in a clinical setting

In the present study we provide a clinical approach to screen for the growing number of students with dyslexia entering higher education programs. After having developed reliable tools to rapidly detect these students, there is a need to select the particular combination of tests that produces the best screening procedure for dyslexia. We therefore took advantage of a supervised machine learning approach, namely conditional inference trees analysis (Horthorn et al., 2006; 2015) to obtain supervised classification and identification of the most relevant tests and associated cut-offs (i.e., conditional cut-off scores) to identify students with dyslexia using an optimal order of test administration. This methodology is very dynamic where a series of different tests may be added to the decision sequence with individual (conditional) cut-offs or else the same test may appear at more than one point in the sequence with different (conditional) cut-offs (or any combination of the two scenarios). The tests and conditional cut-off scores allowing the best discrimination between readers with dyslexia and skilled readers were selected from our study to create this decision tree (see Figures 2 and 3 for detailed results).

To provide an illustration of how these two classification procedures might be used by clinicians, we present two examples below (Participants 1 and 2, see Figure 4 and Table 3). In Table 3, both participants' scores can be inspected together with the respective cut-off scores for the individual tests as computed from the original ROC analysis. Importantly, the ROC cut-off scores can be used to identify the presence of a deficit when the test is used in isolation. However, in the procedure devised using the conditional inference classification methodology, the scores must be used in the sequential order determined by the decision tree. These latter cut-off scores are therefore termed "conditional cut-off scores".

The scores obtained by **Participant 1** (see Table 3) would lead to a final classification of 'dyslexic' using both decision trees. In the first decision tree model (Figure 4.A), the participant would be classified by the ARHQ score which is above the conditional cut-off score (ARHQ > 0.47) and the pseudoword reading score which is below the conditional cut-off score (pseudoword reading ≤ 105). Using the second decision tree (Figure 4.B), Participant 1 would be classified by the Alouette score and then the phonemic awareness score which are both below their respective conditional cut-off scores (Alouette reading score ≤ 419.45; phonemic awareness score (left part) ≤ 1.75).

The scores obtained by **Participant 2** (see Table 3) would also lead to a final classification of 'dyslexic' using both trees albeit via a different route through the first decision tree from Participant 1 (Figure 4.A). In the first tree, the Participant 2 would be classified by the ARHQ score which is below the conditional cut-off score (ARHQ ≤ 0.47) and then using the phonemic awareness conditional cut-off score (≤ 1.06). Using the second decision tree (Figure 4.B), this participant would be classified by the Alouette reading fluency score and the phonemic awareness score which are both below their respective conditional cut-offs (Alouette reading fluency ≤ 419.45; phonemic awareness ≤ 1.75).

Please insert Table 3 about here

Please insert Figure 4.A about here

Please insert Figure 4.B about here

Implications for research and practice

This study provides French-speaking researchers as well as practitioners who support adults with dyslexia, with a battery of seven tests, each of which can reliably screen for dyslexia. This test battery consists of: 1) a targeted history of the reading difficulties encountered during the adult's lifetime with the administration of the ARHQ; 2) an assessment of reading abilities with the Alouette reading fluency test; 3) a two-minute pseudoword reading test focused on decoding abilities; 4) a connected-text reading fluency test; 5) a phonemic awareness test; 6) a spelling test; and 7) a one-minute word reading test. These tests can help researchers to collect most of the information regarding the accepted cognitive markers of dyslexia in adults.

This information can readily be used by practitioners to reliably screen for higher education students with dyslexia and will contribute towards all aspects (categorical, explanatory, action-oriented) of a later diagnostic process (Tops & Brysbaert, 2020) as it includes tests providing reading indicators (word and pseudoword reading, connected-text reading fluency) and reading related skills (phonemic awareness), clearly identified in the literature as persistently deficient in adults with dyslexia. Our results indicated that the two-minute pseudoword reading test, the Alouette and the ARHQ had excellent discriminatory power, although the one-minute word reading and phonemic awareness tests also have good discriminatory power.

Comparison of discriminatory power with assessment batteries from other alphabetic orthographies that have also been based on this approach are hampered even when similar aspects of skill have been examined because of the variation in how these variables have been operationalized. For example, assessments of word reading vary between measures such as single word reading (e.g., DAST, Harrison and Nichols (2005), for English), the Wordchains test, where participants have 3 minutes to mark inter-word spaces in as many triplets as possible (e.g., Wolff and Lundberg (2003), for Swedish), and a one-minute sentence completion test (e.g., Fernandes et al. (2017), for European-Portuguese). Pseudoword reading tests also differ between techniques such as nonsense passage reading, where nonwords are presented within a meaningless text, which is loosely comparable to the method used with words in the Alouette (e.g., DAST, Harrison and Nichols (2005), for English), pseudohomophone choice (e.g., Wolff and Lundberg (2003), for Swedish) and single pseudoword reading (e.g., Mejia et al. (2012), for Spanish). The same holds true for phonemic awareness tests, which vary greatly in their complexity from spoonerism tests (e.g., Nergard-Nilssen and Eklund (2017), for Norwegian) to first phoneme deletion (e.g., Cavalli et al. (2017), for French) or phoneme segmentation (e.g., Mejia et al. (2012), for Spanish).

What is unclear is whether this cross-linguistic variation in the composition of tools reflects underlying differences in the transparency of the orthographic systems or other factors, most notably, theoretical perspectives on the causes of dyslexia but also the range of available assessment tools in each language. In any case, it would seem advisable for the nature of the orthography to be taken more explicitly into consideration in the development of assessment batteries for adults with dyslexia as orthographic transparency not only influences the speed of learning to read (Seymour et al, 2003) but also the manifestations of dyslexia in both children (Saksida et al. 2016) and adults (Paulesu et al., 2001). Further, the transparency of the orthographic system may have an impact on the discriminatory power of the tests. For example, our pseudoword reading test has excellent psychometric discriminative properties, much stronger than that shown by an identical type of test used by Rouweller et al. (2020) but administered in Dutch, an orthographic system considered as more transparent than French (Seymour et al., 2003; see also Schmalz et al, 2015). Moreover, Ré et al. (2011) suggest that in the case of Italian (a highly transparent orthography), the most discriminating tests are those assessing phonological automaticity, which measure speed rather than accuracy. In contrast, results from the York YAAR-R battery for dyslexia screening in higher education in the UK (Warmington et al., 2013) showed that in English (a highly opaque orthography) students with dyslexia presented impairment in text reading rate, spelling rate, reading comprehension and written skills. Additionally, using just literacy measures (i.e., reading time, reading rate, reading accuracy and spelling error rate) resulted to a good discriminatory power with overall classification rate of 93%. This issue deserves further examination in cross-linguistic research as the answers will inform practitioners about which tools are optimal for their particular language context, including which tools are likely to generalise for use within groupings of similar languages.

A further factor for consideration in comparing studies in this area concerns the sex ratio within the assessed groups. As in the present study, previous research on university students has been based on groups where females were in the majority (e.g., Callens et al., 2012; Cavalli et al., 2018; Hatcher et al., 2010; Tops et al., 2012), likely reflecting the greater prevalence of female than male students in university education. Although this accurately represents the student body that university disability services will encounter, epidemiological studies typically demonstrate that males are diagnosed with dyslexia more frequently than females, with a male: female sex ratio ranging from 1.5:1 to 3.3:1 (e.g., Arnett et al., 2017; Di Folco, Guez, Peyre & Ramus, 2022; Quinn & Wagner, 2015). This male: female sex ratio may be explained by greater variance in male reading performance. However, while male performances are overrepresented at the lower tail of the reading distribution, there is however no sex difference at the higher tail of the distribution (Arnett et al., 2017). Interestingly, university students may constitute a special population in which sex differences might be largely reduced or even disappear, partly due to the selection process at university entrance. In support of this hypothesis, a recent study on dyslexia prevalence conducted on 686 second-year Social Sciences university students found no difference in sex ratio (López-Escribano et al., 2018). The sex ratio of the current study (female: male ratio of 1.9:1) may thus be representative of the sex ratio favouring females in higher education (around 60% depending on subject area). Future research should explore whether sex differences exert any influence over the efficacy of screening batteries.

A screening battery must be associated with a reasonable time demand to be efficient for practitioners (and therefore cost-effective) and it must have high specificity for minimising false positives and maximising PPV at this first stage of student access to learning support. Existing batteries vary greatly according to the number of tests they contain. For example, the DAST (Harrison & Nichols, 2005) involves the administration of 11 tests (reading (decoding and fluency skills), spelling, phonemic awareness, motor coordination, auditory abilities, processing speed, non-verbal reasoning, semantic knowledge and two questionnaires on reading difficulties and demographic information). In contrast, the battery by Ré et al. (2011) contains only 2 tests (connected-text reading fluency and spelling under articulatory suppression), and the short form of the YAAR-R battery (Warmington et al., 2013) which contains only 4 tests out of 10 (reading comprehension, written précis, RAN digits and spoonerism) yields poor 75% sensitivity and high 99% specificity.

The conditional inference tree methodology explored in the present study offered a way of moving towards a more efficient screening battery using only a subset of the seven tests examined. For French, two possibilities emerged each with similarly high sensitivity (~87%). One decision tree was based on three tests (ARHQ, phonemic awareness and pseudoword reading) and the other on two tests (Alouette, phonemic awareness), each administered in a standardized order and with designated conditional cut-off scores. More work is needed to investigate how well such methods fit the requirements of practice. Additionally, the patterns of dyslexic performance revealed by the decision trees might also be of interest in further research. For example, the decision tree in Figure 2 initially distinguishes two groups of students with diagnosed dyslexia - those who self-report a history of reading difficulties and those who do not. Deacon et al. (2012) points out that even when a formal diagnosis is not yet present, the profile of the former group is very similar to that of diagnosed dyslexics. The latter group who do not self-report such a severe level of reading difficulty and yet are classified as dyslexic due to poor phonemic awareness at the next decision point may have developed effective compensation strategies or have difficulties that emerged later than the elementary period of schooling surveyed in the ARHQ. Although such patterns remain speculative, future research will establish whether this classification method has the potential to signpost different dyslexic profiles for follow-up investigation in the diagnostic process.

Finally, confirmation of the classification outcome using a full diagnostic procedure is necessary to complete the categorical, explanatory and action-oriented aspects of diagnosis (Tops & Brysbaert, 2020). According to the reported findings, we propose that from the perspective of screening in French, one should use the ARHQ (if available) which explicitly asks about earlier experiences of reading along with a formal case history, followed by a validation of the reading and reading-related deficits using only three tests (i.e., the Alouette reading fluency test, the phonological awareness test, and the pseudoword reading test) which are consistent with the findings that phonological processing difficulties is a lifelong indicator of dyslexia. A complete cognitive assessment should then be proposed, including the assessment of verbal and non-verbal IQ, oral language skills and semantics, reading comprehension (examining both literal and inferential processing), and the efficiency of executive function and memory (see Colé, Cavalli, and Duncan, 2020).

Conclusion

This study reports on the discriminatory power of seven tests relevant for diagnosing dyslexia in French university students together with a methodology outlining how a subset of these tests might best be combined to form short standardised screening procedure with high sensitivity. The behavioural data from our group of university students with dyslexia revealed persistent deficits in reading and in skills associated with success in learning to read. Given the significant costs to academic performances of a failure to diagnose these difficulties, it is important to implement reliable assessment procedures on university entry. At university, there is a need for fast eligibility decisions about exam facilities and other compensatory means. In clinical practice, many practitioners and educational services in higher education agree that the accuracy of the classification procedures for dyslexia can be improved. The present results are therefore of tremendous importance as they provide a short screening protocol of 4 tests with their relative conditional cut-off scores which is a user-friendly method with excellent psychometric discriminative properties for classifying students with dyslexia. Additional assessments may then be recommended to highlight not only deficits but also cognitive strengths for students with dyslexia in order that the appropriate adjustments and accommodations can be put in place to support their studies.

Open Practices Statement

The materials used in this study are available on the *Open Science Framework* website *[website hidden for review]*, and this study was not preregistered.

References

Alves, R., & Castro, S. L. (2005). Despistagem da dislexia em adultos através do questionário história de leitura [Adult dyslexia screening using a portuguese self-report measure].

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub. https://doi.org/10.1176/appi.books.9780890425596

American Psychiatric Association (2000). Diagnostic and Statistical Manual of Mental Disorders Fourth Edition Text Revision (DSM-IV-TR). Washington DC: American Psychiatric Association. https://doi.org/10.1176/appi.books.9780890423349

Arnett, A.B., Pennington, B.F., Peterson, R.L., Willcutt, E.G., DeFries, J.C., & Olson, R.K. (2017). Explaining the sex difference in dyslexia. *Journal of Child Psychology and Psychiatry*, 58(6), 719–727. https://doi.org/10.1111/jcpp.12691

Bergey, B. W., Deacon, S. H., & Parrila, R. K. (2017). Metacognitive reading and study strategies and academic achievement of university students with and without a history of reading difficulties. *Journal of Learning Disabilities*, *50*(1), 81-94. https://doi.org/10.1177/0022219415597020

Bertrand, D., Fluss, J., Billard, C., & Ziegler, J. C. (2010). Efficacité, sensibilité, spécificité : Comparaison de différents tests de lecture. *L'Année Psychologique*, *110*(2), 299-320. https://doi.org/10.4074/S000350331000206X

Bjornsdottir, G., Halldorsson, J. G., Steinberg, S., Hansdottir, I., Kristjansson, K., Stefansson, H., & Stefansson, K. (2014). The adult reading history questionnaire (ARHQ) in Icelandic: psychometric properties and factor structure. *Journal of Learning Disabilities*, *47*(6), 532-542. https://doi.org/10.1177/0022219413478662

Blampain, E., Gosse, C., & Van Reybroeck, M. (2021). Copying skills in children with and without dyslexia. *Reading and Writing*, *34*(4), 859-885. https://doi.org/10.1007/s11145-020-10094-6

Bonin, P., Méot, A., & Bugaiska, A. (2018). Concreteness norms for 1,659 French words: Relationships with other psycholinguistic variables and word recognition times. Behavior research methods, 50(6), 2366–2387. https://doi.org/10.3758/s13428-018-1014-y

Bradshaw, A. R., Woodhead, Z. V., Thompson, P. A., & Bishop, D. V. (2021). Profile of language abilities in a sample of adults with developmental disorders. *Dyslexia*. https://doi.org/10.1002/dys.1672

Brèthes H, **Cavalli E**, Denis-Noël A, Melmi J-B, El Ahmadi A, Bianco M and Colé P (2022) Text Reading Fluency and Text Reading Comprehension Do Not Rely on the Same Abilities in University Students With and Without Dyslexia. *Front. Psychol.* 13:866543. doi: 10.3389/fpsyg.2022.866543

Broggi, M., Ready, R. E., & Moore, D. L. (2019). Screening for reading disability in university students with phonological processing and working memory tasks. *Dyslexia*, *25*(3), 256-266. https://doi.org/10.1002/dys.1625

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*(1), 5-13.

Callens, M., Tops, W., & Brysbaert, M. (2012). Cognitive profile of students who enter higher education with an indication of dyslexia. *PloS One*, *7*(6), e38081. https://doi.org/10.1371/journal.pone.0038081

Cavalli, E., Casalis, S., El Ahmadi, A., Zira, M., Poracchia-George, F., & Colé, P. (2016). Vocabulary skills are well developed in university students with dyslexia: Evidence from multiple case studies. *Research in Developmental Disabilities*, *51*, 89-102. https://doi.org/10.1016/j.ridd.2016.01.006

Cavalli, E., Colé, P., Brèthes, H., Lefevre, E., Lascombe, S., & Velay, J. L. (2019). E-book reading hinders aspects of long-text comprehension for adults with dyslexia. *Annals of Dyslexia*, *69*(2), 243-259. https://doi.org/10.1007/s11881-019-00182-w

Cavalli, E., Colé, P., Leloup, G., Poracchia-George, F., Sprenger-Charolles, L., & El Ahmadi, A. (2018). Screening for dyslexia in French-Speaking University students: An evaluation of the detection accuracy of the Alouette test. *Journal of Learning Disabilities*, *51*(3), 268-282. https://doi.org/10.1177/0022219417704637

Cavalli, E., Duncan, L. G., Elbro, C., El Ahmadi, A., & Colé, P. (2017). Phonemic—Morphemic dissociation in university students with dyslexia: An index of reading compensation? *Annals of Dyslexia*, *67*(1), 63-84. https://doi.org/10.1007/s11881-016-0138-y

Chetail, F., Porteous, F., Patout, P. A., Content, A., & Collette, E. (2019). BOQS: une échelle de qualité orthographique pour l'adulte. *Revue de Neuropsychologie*, *11*(3), 214-226. https://doi.org/10.1684/nrp.2019.0518

Chevalier, T. M., Parrila, R., Ritchie, K. C., & Deacon, S. H. (2017). The role of metacognitive reading strategies, metacognitive study and learning strategies, and behavioral study and learning strategies in predicting academic success in students with and without a history of reading difficulties. *Journal of Learning Disabilities*, *50*(1), 34-48. https://doi.org/10.1177/0022219415588850

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* (2nd ed.), Erlbaum, Hillsdale, NJ.

Colé, P., Duncan, L. G., & Cavalli, E. (2020). Les compensations de l'adulte dyslexique de niveau universitaire [Compensations of university students with dyslexia]. In P. Colé, E.

Cavalli & L. G. Duncan (Eds.), *La dyslexie à l'âge adulte : Approche neuropsychologique* (pp. 287-324). De Boeck Supérieur.

Couvignou, M., Peretz, I., & Ramus, F. (2019). Comorbidity and cognitive overlap between developmental dyslexia and congenital amusia. *Cognitive Neuropsychology*, *36*(1-2), 1-17. https://doi.org/10.1080/02643294.2019.1578205

Cutting, L.E., & Scarborough, H.S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10(3), 277-299. https://doi.org/10.1207/s1532799xssr1003_5

Deacon, S. H., Cook, K., & Parrila, R. (2012). Identifying high-functioning dyslexics: is self-report of early reading problems enough? *Annals of Dyslexia*, *62*(2), 120-134. https://doi.org/10.1007/s11881-012-0068-2

Deacon, S. H., Tucker, R., Bergey, B. W., Laroche, A., & Parrila, R. (2017). Personalized outreach to university students with a history of reading difficulties: Early screening and outreach to support academically at-risk students. *Journal of College Student Development*, *58*(3), 432-450. https://doi.org/10.1353/csd.2017.0032

De La Haye, F., Gombert, J.-E., Rivière, J., & Rocher, T. (2008). Les évaluations en lecture dans le cadre de la journée d'appel de préparation à la défense [Reading assessments during the citizenship day], année 2007. *Les notes d'information D.E.P.P.*

Del Tufo, S. N., & Earle, F. S. (2020). Skill profiles of college students with a history of developmental language disorder and developmental dyslexia. *Journal of Learning Disabilities*, *53*(3), 228-240. https://doi.org/10.1177/0022219420904348

Depessemier, P., & Andries, C. (2009). *GL&SCHR: dyslexie bij+ 16-jarigen: test voor Gevorderd Lezen & SCHRijven*. Garant.

Di Folco, C., Guez, A., Peyre, H., & Ramus, F. (2022). Epidemiology of reading disability: A comparison of DSM-5 and ICD-11 criteria, *Scientific Studies of Reading*, 26(4), 337-355, https://doi.org/10.1080/10888438.2021.1998067

Dunn, L. M., Thériault-Whalen, C. M., & Dunn, L. M. (1993). Echelles de vocabulaire en image Peabody. Adaptation française du Peabody Picture Vocabulary Test-Revised. Psycan Toronto.

Dunn, M., & Dunn, L. M. (1981). Peabody picture vocabulary test – Revised. Circle Pines, MN: AGS.

Elbro, C., Nielsen, I., & Petersen, D. K. (1994). Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representation of lexical items. *Annals of Dyslexia*, *44*(1), 203-226. https://doi.org/10.1007/BF02648162

Erskine, J. M., & Seymour, P. H. K. (2005). Proximal analysis of developmental dyslexia in adulthood: The Cognitive Mosaic model. *Journal of Educational Psychology*, 97(3), 406–424. https://doi.org/10.1037/0022-0663.97.3.406Fernandes, T., Araújo, S., Sucena, A., Reis, A., & Castro, S. L. (2017). The 1-min screening test for reading problems in college students: Psychometric properties of the 1-min TIL. *Dyslexia*, *23*(1), 66-87. https://doi.org/10.1002/dys.1548

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239-256. https://doi.org/10.1207/S1532799XSSR0503_3

Georgiou, G. K., Martinez, D., Vieira, A. P. A., & Guo, K. (2021). Is orthographic knowledge a strength or a weakness in individuals with dyslexia? Evidence from a meta-analysis. *Annals of Dyslexia*, *71*(1), 5-27. https://doi.org/10.1007/s11881-021-00220-6

Gola-Asmussen, C., Lequette, C., Pouget, G., Rouyet, C. et Zorman, M. (2011). *ECLA 16+ : Évaluation des compétences de lecture chez l'adulte de plus de 16 ans* [Reading skills assessment in adults starting from 16 years-old]. Grenoble : Université de Provence Aix-Marseille I - Cognisciences LSE Université Pierre Mendès.

Gosse, C., & Van Reybroeck, M. (2020). Do children with dyslexia present a handwriting deficit? Impact of word orthographic and graphic complexity on handwriting and spelling performance. *Research in Developmental Disabilities*, *97*, 103553. https://doi.org/10.1016/j.ridd.2019.103553

Grenier, B. (1999). *Évaluation de la décision médicale : Introduction à l'analyse médico-économique [Medical decision assessment : Introduction to medico-economic analysis]*. Paris: Masson.

Haft, S. L., Myers, C. A., & Hoeft, F. (2016). Socio-emotional and cognitive resilience in children with reading disabilities. *Current Opinion in Behavioral Sciences*, *10*, 133-141. https://doi.org/10.1016/j.cobeha.2016.06.005

Harrison, A. G., & Nichols, E. (2005). A validation of the Dyslexia Adult Screening Test (DAST) in a post-secondary population. *Journal of Research in Reading*, *28*(4), 423-434. https://doi.org/10.1111/j.1467-9817.2005.00280.x

Hatcher, J., Snowling, M. J., & Griffiths, Y. M. (2002). Cognitive assessment of dyslexic students in higher education. *British Journal of Educational Psychology*, *72*(1), 119-133. https://doi.org/10.1348/000709902158801

Hothorn, T., & Zeileis, A. (2015). partykit: A modular toolkit for recursive partytioning in R. *The Journal of Machine Learning Research*, *16*(1), 3905-3909.

Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2015). party: a laboratory for recursive partytioning. R package version 1.2-3.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics*, *15*(3), 651–674. https://doi.org/10.1198/106186006X133933

Hang, Y., Fong, Y., Wei, J., and Feng, Z. (2011). Borrowing information across populations in estimating positive and negative predictive values. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **60**, 633–653. https://doi.org/10.1111/j.1467-9876.2011.00761.x

Jung, Y., & Hu, J.(2015). A *K*-fold averaging cross-validation procedure, Journal of Nonparametric Statistics, 27:2, 167-179, DOI: 10.1080/10485252.2015.1010532)

Keenan, J.M, Betjemann, R.S., & Olson, R.K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12(3), 281-300. https://doi.org/10.1080/10888430802132279

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... & Benesty, M. (2020). Package 'caret'. *The R Journal*, 223.

Law, J. M., Vandermosten, M., Ghesquiere, P., & Wouters, J. (2014). The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. *Frontiers in Human Neuroscience*, *8*, 482. https://doi.org/10.3389/fnhum.2014.00482

Leblanc, M. (1913). L'Écharpe de Soie rouge [The red silk scarf]. In *Les confidences d'Arsène Lupin* (Editions P). Paris.

Lefavrais, P. (1967). *Test de l'Alouette* [Alouette Test].

Lefavrais, P. (2005). *Test de l'Alouette* (ré-étalonné en 2005) [Alouette Test (2005 standardization)]. ECPA. Paris.

Lefly, D. L., & Pennington, B. F. (2000). Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*, *33*(3), 286-296. https://doi.org/10.1177/002221940003300306

Levshina, N. (2015). Conditional inference trees and random forests. In How to do linguistics with R (pp. 291–300). Amsterdam: John Benjamins.

Lindgrén, S. A., & Laine, M. (2007). The adaptation of an adult group screening test for dyslexia into Finland-Swedish: Normative data for university students and the effects of language background on test performance. *Scandinavian Journal of Psychology*, *48*(5), 419-432. https://doi.org/10.1111/j.1467-9450.2007.00592.x

López-Escribano, C., Suro Sánchez, J., & Leal Carretero, F. (2018). Prevalence of developmental dyslexia in Spanish University Students. *Brain Sciences*, *8*(5), 82. https://doi.org/10.3390/brainsci8050082

Lüdtke, J., Froehlich, E., Jacobs, A. M., & Hutzler, F. (2019). The SLS-Berlin: validation of a German computer-based screening test to measure reading proficiency in early and late adulthood. *Frontiers in Psychology*, *10*, 1682. https://doi.org/10.3389/fpsyg.2019.01682

Lyon, G. R. (1995). Toward a definition of dyslexia. *Annals of Dyslexia*, *45*(1), 1-27.

Macdonald, S. J., & Deacon, L. (2019). Twice upon a time: Examining the effect socio-economic status has on the experience of dyslexia in the United Kingdom. *Dyslexia*, *25*(1), 3-19. https://doi.org/10.1002/dys.1606

Marchetti, R., Pinto, S., Spieser, L., Vaugoyeau, M., Cavalli, E., El Ahmadi, A., Assaiante, C., & Colé, P. (2023). Phoneme Representation and Articulatory Impairment: Insights from Adults with Comorbid Motor Coordination Disorder and Dyslexia. *Brain sciences*, *13*(2), 210. https://doi.org/10.3390/brainsci13020210

Martin, J., Colé, P., Leuwers, C., Casalis, S., Zorman, M., & Sprenger-Charolles, L. (2010). Reading in French-speaking adults with dyslexia. *Annals of Dyslexia*, *60*(2), 238-264. https://doi.org/10.1007/s11881-010-0043-8

Martin, J., Frauenfelder, U. H., & Cole, P. (2014). Morphological awareness in dyslexic university students. *Applied Psycholinguistics*, *35*(6), 1213-1233. https://doi.org/10.1017/S0142716413000167

Maurer, U. (2003). *Functional brain mapping of development and familial risk for dyslexia in kindergarten children* (Doctoral dissertation, University of Zurich).

Mejía, C., Diaz, A., Jiménez, J. E., & Fabregat, R. (2012). BEDA: a computerized assessment battery for dyslexia in adults. *Procedia-Social and Behavioral Sciences*, *46*, 1795-1800. https://doi.org/10.1016/j.sbspro.2012.05.381

Miller-Shaul, S. (2005). The characteristics of young and adult dyslexics readers on reading and reading related cognitive tasks as compared to normal readers. *Dyslexia*, *11*(2), 132-151. https://doi.org/10.1002/dys.290

Mostaert, C., Smits, I., Tops, W., De Kerf, L., Liekens, E., & Schraeyen, K. (2016). *Test geautomatiseerd spellen voor jongvolwassenen* [Test for Automatized Spelling for young adults].

Nergård-Nilssen, T., & Eklund, K. (2018). Evaluation of the psychometric properties of "the Norwegian screening test for dyslexia". *Dyslexia*, *24*(3), 250-262. https://doi.org/10.1002/dys.1577

Nergård-Nilssen, T., & Hulme, C. (2014). Developmental dyslexia in adults: Behavioural manifestations and cognitive correlates. *Dyslexia*, *20*(3), 191-207. https://doi.org/10.1002/dys.1477

Nicolson, R. I., & Fawcett, A. J. (1997). Development of objective procedures for screening and assessment of dyslexic students in higher education. *Journal of Research in Reading*. https://doi.org/10.1111/1467-9817.00022

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™ [A lexical database for contemporary french: LEXIQUE™]. *L'Année Psychologique*, *101*(3), 447-462. https://doi.org/10.3406/psy.2001.1341

Obidziński, M. (2020). Response frequencies in the conjoint recognition memory task as predictors of developmental dyslexia diagnosis: A decision-trees approach. *Dyslexia*. https://doi.org/10.1002/dys.1655

Parrila, R., Georgiou, G., & Corkett, J. (2007). University students with a significant history of reading difficulties: What is and is not compensated? *Exceptionality Education International*, *17*(2), 195-220. https://doi.org/10.5206/eei.v17i2.7604

Paulesu, E., Démonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., ... & Frith, U. (2001). Dyslexia: Cultural diversity and biological unity. *Science*, *291*(5511), 2165-2167. https://doi.org/10.1126/science.1057179

Peereman, R., Lété, B., & Sprenger-Charolles, L. (2007). Manulex-infra: Distributional characteristics of grapheme—phoneme mappings, and infralexical and lexical units in child-directed written material. *Behavior Research Methods*, *39*(3), 579-589. https://doi.org/10.3758/BF03193029

Peereman, R., Sprenger-Charolles, L., & Messaoud-Galusi, S. (2013). The contribution of morphology to the consistency of spelling-to-sound relations: A quantitative analysis based on French elementary school readers. *L'Année Psychologique*, *113*(1), 3-33. https://doi.org/10.4074/S0003503313001012

Peter, B., Albert, A., & Gray, S. (2020). Spelling errors reveal underlying sequential and spatial processing deficits in adults with dyslexia. *Clinical Linguistics & Phonetics*, 1-30. https://doi.org/10.1080/02699206.2020.1780322

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*, 283-307. https://doi.org/10.1146/annurev-clinpsy-032814-112842

Quinn, J. M., & Wagner, R. K. (2015). Gender differences in reading impairment and in the identification of impaired readers: Results from a large-scale study of at-risk readers. *Journal of Learning Disabilities*, 48(4), 433–445. https://doi.org/10.1177/0022219413508323

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. https://doi.org/10.1080/02643294.2012.677420

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia: insights from a multiple case study of dyslexic adults. *Brain*, *126*(4), 841-865. https://doi.org/10.1093/brain/awg076

Ransby, M. J., & Lee Swanson, H. (2003). Reading comprehension skills of young adults with childhood diagnoses of dyslexia. *Journal of Learning Disabilities*, *36*(6), 538-555. https://doi.org/10.1177/00222194030360060501

Raven, J. C., Court, J. H., & Raven, J. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General overviews*. Oxford, England: Oxford Psychologists Press.

Re, A. M., Tressoldi, P. E., Cornoldi, C., & Lucangeli, D. (2011). Which tasks best discriminate between dyslexic university students and controls in a transparent language? *Dyslexia*, *17*(3), 227-241. https://doi.org/10.1002/dys.431

Reis, A., Araújo, S., Morais, I. S., & Faísca, L. (2020). Reading and reading-related skills in adults with dyslexia from different orthographic systems: a review and meta-analysis. *Annals of Dyslexia*, *70*(3), 339-368. https://doi.org/10.1007/s11881-020-00205-x

Reynolds, A. E., & Caravolas, M. (2016). Evaluation of the Bangor Dyslexia Test (BDT) for use with Adults. *Dyslexia*, *22*(1), 27-46. https://doi.org/10.1002/dys.1520

Richardson, J. T., & Wydell, T. N. (2003). The representation and attainment of students with dyslexia in UK higher education. *Reading and Writing*, *16*(5), 475-503. https://doi.org/10.1023/A:1024261927214

Ritchie, L. J., & Tuokko, H. (2011). Clinical decision trees for predicting conversion from cognitive impairment no dementia (CIND) to dementia in a longitudinal population-based study. *Archives of clinical neuropsychology, 26(1), 16-25.*

Rose, L. T., & Rouhani, P. (2012). Influence of verbal working memory depends on vocabulary: Oral reading fluency in adolescents with dyslexia. *Mind, Brain, and Education*, *6*(1), 1-9. https://doi.org/10.1111/j.1751-228X.2011.01135.x

Rouweler, L., Varkevisser, N., Brysbaert, M., Maassen, B., & Tops, W. (2020). The Flamingo test: a new diagnostic instrument for dyslexia in Dutch higher education students. *European Journal of Special Needs Education*, *35*(4), 529-543. https://doi.org/10.1080/08856257.2019.1709703

Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J. F., Bricout, L., ... & Ramus, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology*, *52*(10), 1503. https://doi.org/10.1037/dev0000184

Sarda-Espinosa, A., Subbiah, S., & Bartz-Beielstein, T. (2017). Conditional inference trees for knowledge extraction from motor health condition data. *Engineering Applications of Artificial Intelligence*, *62*, 26-37.

Seymour, P. H., Aro, M. & Erskine, J. M., (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143-174. https://doi.org/10.1348/000712603321661859

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review*, 22(6), 1614–1629. https://doi.org/10.3758/s13423-015-0835-2

Schraeyen, K., Vanderauwera, J., & Vandermosten, M., (2020). Les compétences de traitement phonologique chez les adultes avec une dyslexie [Phonological processing skills in adults with dyslexia]. In P. Colé, E. Cavalli & L. G. Duncan (Eds.), *La dyslexie à l'âge adulte: Approche neuropsychologique* (pp. 40-64). De Boeck Supérieur.

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls: Results of the Connecticut Longitudinal Study. *Jama*, *264*(8), 998-1002. https://doi.org/10.1001/jama.1990.03450080084036

Singleton, C., Horne, J., & Simmons, F. (2009). Computerised screening for dyslexia in adults. *Journal of Research in Reading*, *32*(1), 137-152. https://doi.org/10.1111/j.1467-9817.2008.01386.x

Snowling, M. J. (2000). Foundations of reading acquisition and dyslexia: Implications for early intervention. *British Journal of Educational Psychology*, *70*, 275.

Sprenger-Charolles, L. (2019). Developmental Dyslexia in French. In L. Verhoeven, C. Perfetti & K. Pugh (Eds.), *Developmental Dyslexia across Languages and Writing Systems* (pp. 50 - 72). Cambridge University Press. https://doi.org/10.1017/9781108553377

Strobl, C., Malley, J., & Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychological Methods*, *14*(4), 323. https://doi.org/10.1037/a0016973

Suárez-Coalla, P., & Cuetos, F. (2015). Reading difficulties in Spanish adults with dyslexia. *Annals of Dyslexia*, *65*(1), 33-51. https://doi.org/10.1007/s11881-015-0101-3

Swanson, H. L. (2012). Adults with reading disabilities: Converting a meta-analysis to practice. *Journal of Learning Disabilities*, *45*(1), 17-30. https://doi.org/10.1177/0022219411426856

Swanson, H. L., & Hsieh, C. J. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research*, *79*(4), 1362-1390. https://doi.org/10.3102/0034654309350931

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*(4857), 1285-1293.

Ter Huurne, D., Ramakers, I., Possemis, N., Banning, L., Gruters, A., Van Asbroeck, S., ... & de Vugt, M. (2023). The Accuracy of Speech and Linguistic Analysis in Early Diagnostics of

Neurocognitive Disorders in a Memory Clinic Setting. *Archives of Clinical Neuropsychology, acac105.*

Tops, W., & Brysbaert, M. (2020). Le dépistage et diagnostic de la dyslexie à l'université [Screening and diagnostic of dyslexia in university]. In P. Colé, E. Cavalli & L. G. Duncan (Eds.), *La dyslexie à l'âge adulte : Approche neuropsychologique* (pp. 202-217). De Boeck Supérieur.

Tops, W., Callens, M., Lammertyn, J., Van Hees, V., & Brysbaert, M. (2012). Identifying students with dyslexia in higher education. *Annals of Dyslexia*, *62*(3), 186-203. https://doi.org/10.1007/s11881-012-0072-6

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2-40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Warmington, M., Stothard, S. E., & Snowling, M. J. (2013). Assessing dyslexia in higher education: The York adult assessment battery-revised. *Journal of Research in Special Educational Needs*, *13*(1), 48-56. https://doi.org/10.1111/j.1471-3802.2012.01264.x

Wolff, U., & Lundberg, I. (2002). The prevalence of dyslexia among art students. *Dyslexia*, *8*(1), 34-42. https://doi.org/10.1002/dys.211

Wolff, U., & Lundberg, I. (2003). A technique for group screening of dyslexia among adults. *Annals of Dyslexia*, *53*(1), 324-339. https://doi.org/10.1007/s11881-003-0015-3

Table and Figure captions

Table 1. Performances (mean (M), standard deviation (SD) and effect size) obtained by skilled readers (SR; n = 65) and readers with dyslexia (DYS; n = 60) on all administered tests. With the exception of the ARHQ, all measures are efficiency scores.

Table 2. Classification based on Receiver Operator Characteristic analysis (diagnosed as readers with and without dyslexia) for all the variable cut-off scores. (Se = sensitivity; Sp = specificity; AUC = area under the curve).

Table 3. Performance of Participants 1 and 2 on the screening tests with associated ROC cut-off scores.

Figure 1. A. Receiver Operator Characteristic (ROC) curves with 95% confidence intervals (left) and optimal cut-off plots (right) for word reading, pseudoword reading, phonemic awareness, and spelling. On the right plot, SR codes for Skilled Readers and DYS for Dyslexic Readers. The horizontal line represents the optimal cut-off value and the associated sensitivity and specificity indicators.

Figure 1. B. Receiver Operator Characteristic (ROC) curves with 95% confidence intervals (left) and optimal cut-off plots (right) for connected-text reading fluency, Alouette reading fluency, and ARHQ tests. On the right plot, SR codes for Skilled Readers and DYS for Dyslexic Readers. The horizontal line represents the optimal cut-off value and the associated sensitivity and specificity indicators.

Figure 2. Conditional inference tree allowing categorizing individuals as readers with or without dyslexia. Complete data from the 7 assessment tests for 125 students (65 skilled readers (SR), 60 dyslexic readers (DYS)) were used for prognostic modelling (word reading, pseudoword reading, phonemic awareness, spelling, connected-text reading fluency, Alouette reading, and ARHQ). For each inner node, the Bonferroni-adjusted *P*-values are given. The proportion of readers with and without dyslexia is displayed for each terminal node in the observed part. In the terminal node participants in the grey cells are misclassified by the algorithm in respect of their initial diagnostic.

Figure 3. Conditional inference tree allowing categorizing individuals as readers with or without dyslexia. Complete data from 6 assessment tests for 125 students (65 skilled readers (SR), 60 dyslexic readers (DYS)) were used for prognostic modelling (word reading, pseudoword reading, phonemic awareness, spelling, connected-text reading fluency, and Alouette reading). For each inner node, the Bonferroni-adjusted *P*-values are given. The proportion of readers with and without dyslexia is displayed for each terminal node in the observed part. In the terminal node participants in the grey cells are misclassified by the algorithm in respect of their initial diagnostic.

Figure 4. A. Illustration of how the classification procedures (with Tree 1) might be used for screening by clinicians. The Figure presents two examples (Participant 1, top part; Participant 2, bottom part). In each graph, bold arrows represent the decision path with respect to the obtained scores (see Table 3) and the conditional cut-off scores.

Figure 4. B. Illustration of how the classification procedures (with Tree 2) might be used for screening by clinicians. The Figure presents two examples (Participant 1, top part; Participant 2, bottom part). In each graph, bold arrows represent the decision path with respect to the obtained scores (see Table 3) and the conditional cut-off scores.

Tables and Figures

Table 1. Performances (mean (M), standard deviation (SD) and effect size) obtained by skilled readers (SR; n = 65) and readers with dyslexia (DYS; n = 60) on all administered tests. With the exception of the ARHQ, all measures are efficiency scores.

	SR group M (SD)	DYS group M (SD)	(t-test) *p* value	Effect Size (Cohen's d)
Word reading	104.8 (22.1)	77.2 (20.9)	(7.1) ***	1.3
Pseudoword reading	135.2 (31.6)	73.8 (29.2)	(11.5) ***	2.0
Phonemic awareness	2.0 (0.4)	1.1 (0.4)	(10.1) ***	2.2
Spelling	75.0 (3)	68.6 (12.2)	(7.8) ***	0.7
Connected-text reading fluency	196.7 (33.2)	140.5 (29.9)	(10.2) ***	1.8
Alouette reading test	513.4 (82.1)	353.9 (74.6)	(12.3) ***	2.0
ARHQ	0.31 (0.10)	0.59 (0.09)	(5.3) ***	2.9

***p < .001

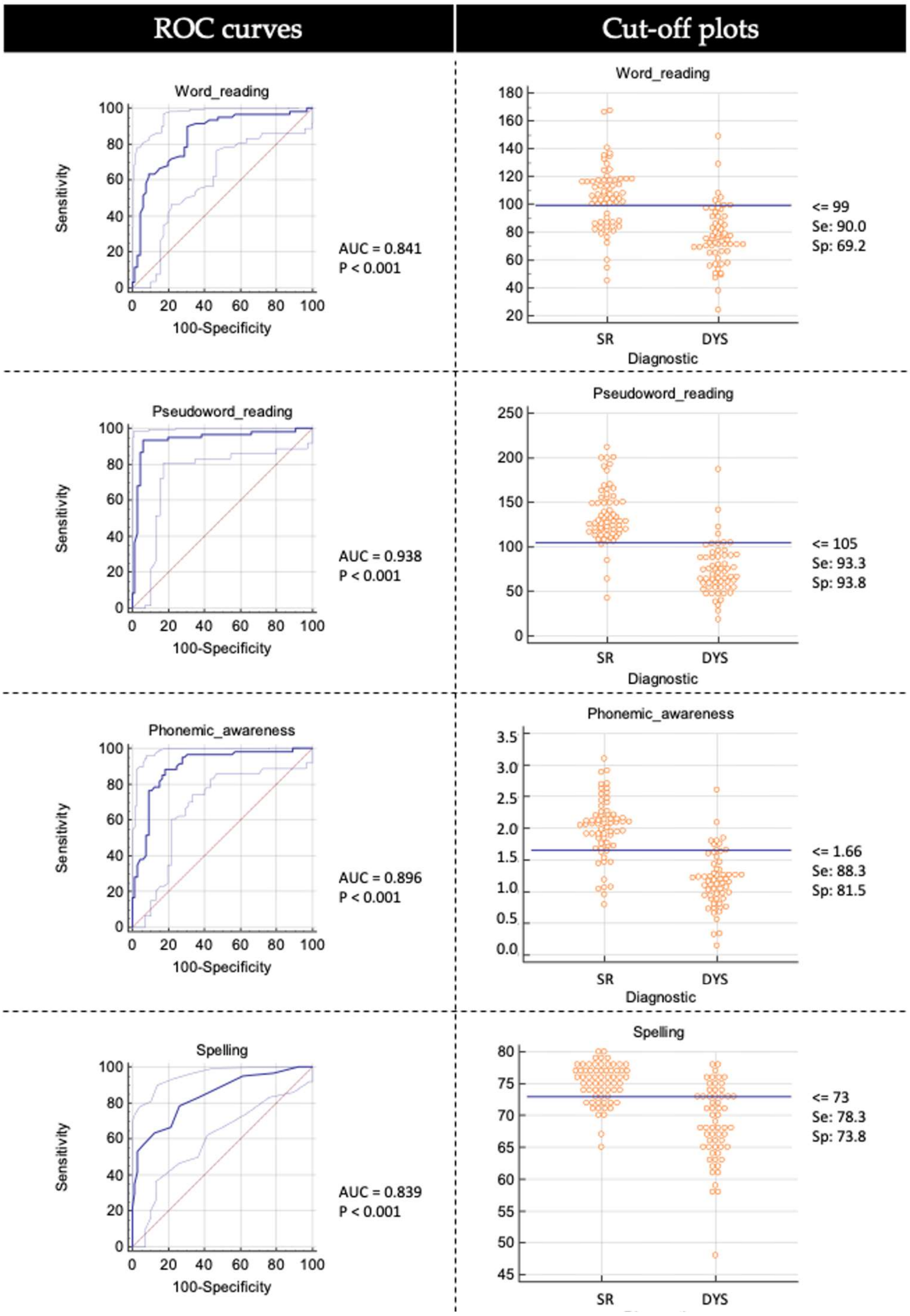

Figure 1. A. Receiver Operator Characteristic (ROC) curves with 95% confidence intervals (left) and optimal cut-off plots (right) for word reading, pseudoword reading, phonemic awareness, and spelling. On the right plot, SR codes for Skilled Readers and DYS for Dyslexic Readers. The horizontal line represents the optimal cut-off value and the associated sensitivity and specificity indicators.

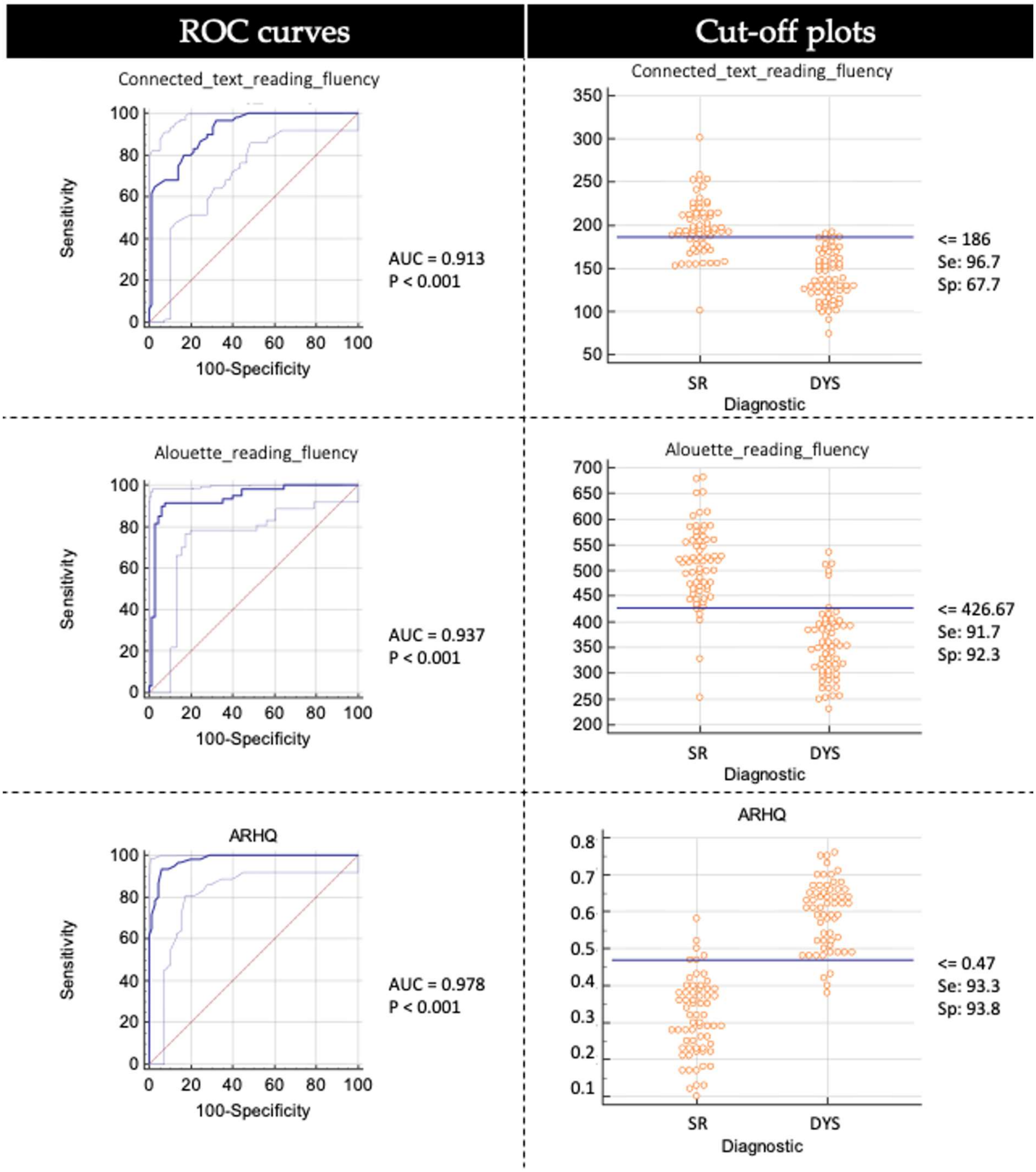

Figure 1. B. Receiver Operator Characteristic (ROC) curves with 95% confidence intervals (left) and optimal cut-off plots (right) for connected-text reading fluency, Alouette reading fluency, and ARHQ tests. On the right plot, SR codes for Skilled Readers and DYS for Dyslexic Readers. The horizontal line represents the optimal cut-off value and the associated sensitivity and specificity indicators.

Table 2. Classification based on Receiver Operator Characteristic analysis (diagnosed as readers with and without dyslexia) for all the variable cut-off scores. (Se = sensitivity; Sp = specificity; AUC = area under the curve)

	Receiver operator characteristic (ROC) curves analysis							
	Cut-off score	AUC (SD)	95% CI	p value	Se	95% CI	Sp	95% CI
Word reading	99	.84 (.03)	[.76 - .90]	***	90.0	[79.5 – 96.2]	69.2	[56.6 – 80.1]
Pseudoword reading	105	.93 (.02)	[.88 - .97]	***	93.3	[83.8 – 98.2]	93.8	[85.0 – 98.3]
Phonemic awareness	1.6	.89 (.02)	[.82 - .94]	***	88.3	[77.4 – 95.2]	81.5	[70.0 – 90.1]
Spelling	73	.83 (.03)	[.76 - .89]	***	78.3	[65.8 – 87.9]	73.8	[61.5 – 84.0]
Connected-text reading fluency	186	.91 (.02)	[.84 - .95]	***	96.7	[88.5 – 99.6]	67.7	[54.9 – 78.8]
Alouette reading test	426.6	.93 (.02)	[.87 - .97]	***	91.6	[81.6 – 97.2]	92.3	[83.0 – 97.5]
ARHQ	0.47	.97 (.00)	[.93 - .99]	***	93.3	[83.8 – 98.2]	93.8	[85.0 – 98.3]

***$p \leq .001$.

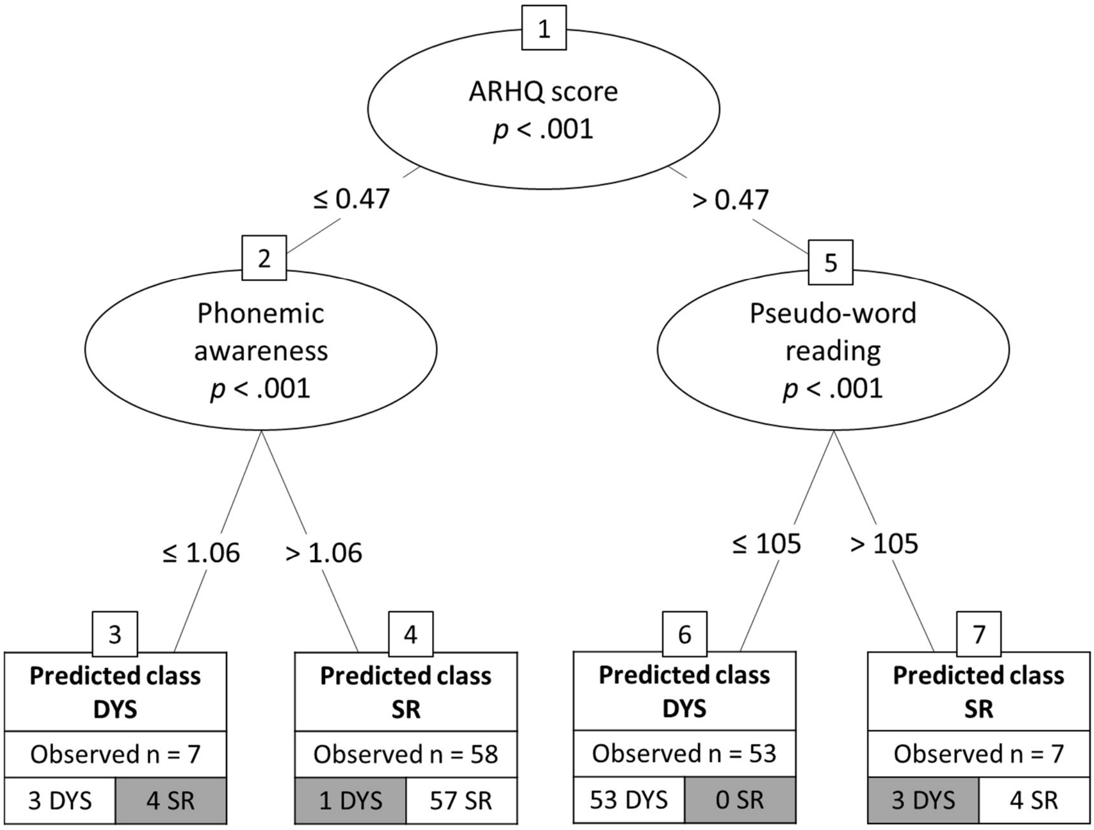

Figure 2. Conditional inference tree allowing categorizing individuals as readers with or without dyslexia. Complete data from the 7 assessment tests for 125 students (65 skilled readers (SR), 60 dyslexic readers (DYS)) were used for prognostic modelling (word reading, pseudoword reading, phonemic awareness, spelling, connected-text reading fluency, Alouette reading, and ARHQ). For each inner node, the Bonferroni-adjusted *P*-values are given. The proportion of readers with and without dyslexia is displayed for each terminal node in the observed part. In the terminal node participants in the grey cells are misclassified by the algorithm in respect of their initial diagnostic.

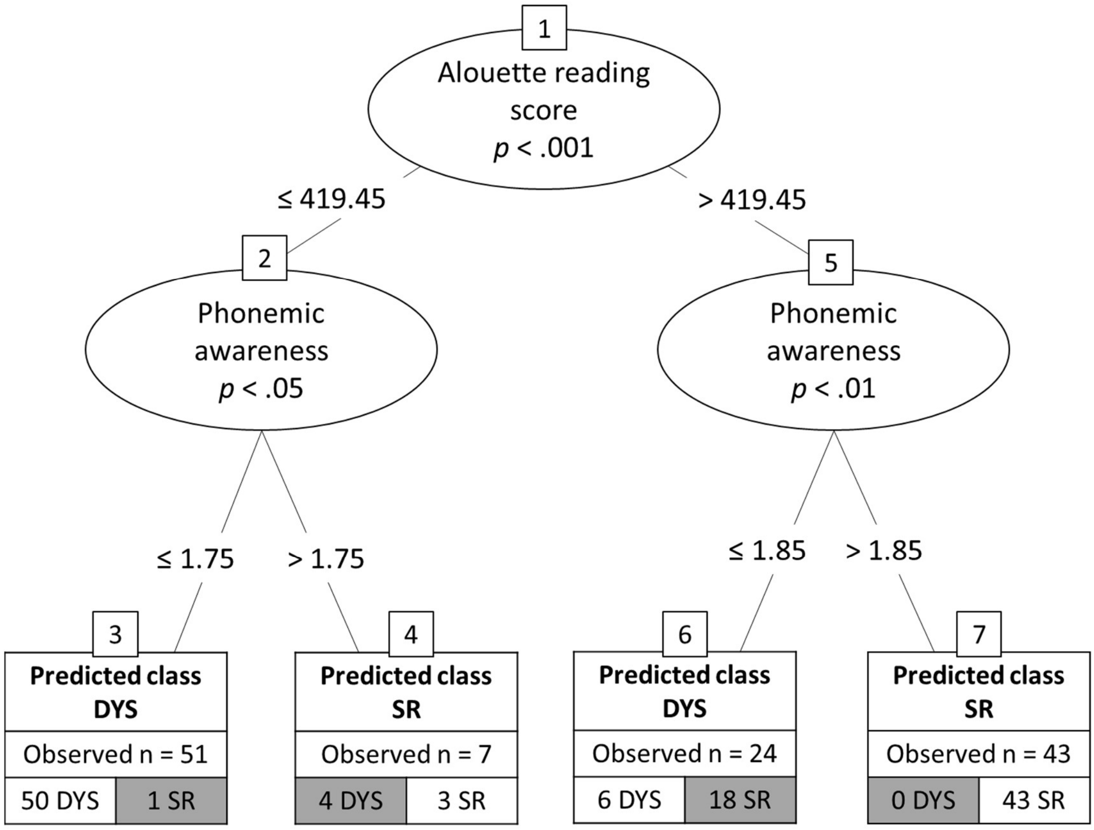

Figure 3. Conditional inference tree allowing categorizing individuals as readers with or without dyslexia. Complete data from 6 assessment tests for 125 students (65 skilled readers (SR), 60 dyslexic readers (DYS)) were used for prognostic modelling (word reading, pseudoword reading, phonemic awareness, spelling, connected-text reading fluency, and Alouette reading). For each inner node, the Bonferroni-adjusted *P*-values are given. The proportion of readers with and without dyslexia is displayed for each terminal node in the observed part. In the terminal node participants in the grey cells are misclassified by the algorithm in respect of their initial diagnostic.

Table 3. Performance of Participants 1 and 2 on the screening tests with associated ROC cut-off scores.

Screening tests	Participant 1	Participant 2	ROC cut-off score
Word reading	**24**	**94**	99
Pseudoword reading	**47**	122.35	105
Phonemic awareness	**1.46**	**0.74**	1.6
Spelling	78	**65**	73
Connected-text reading fluency	**74**	**136**	186
Alouette reading test	**253**	**351.36**	426.6
ARHQ	**0.64**	0.43	0.47

Note: Scores presented in bold are considered as impaired with the use of the ROC cut-off score. Please refer to Figures 4 that presents the conditional cut-off scores computed from the classification tree analysis.

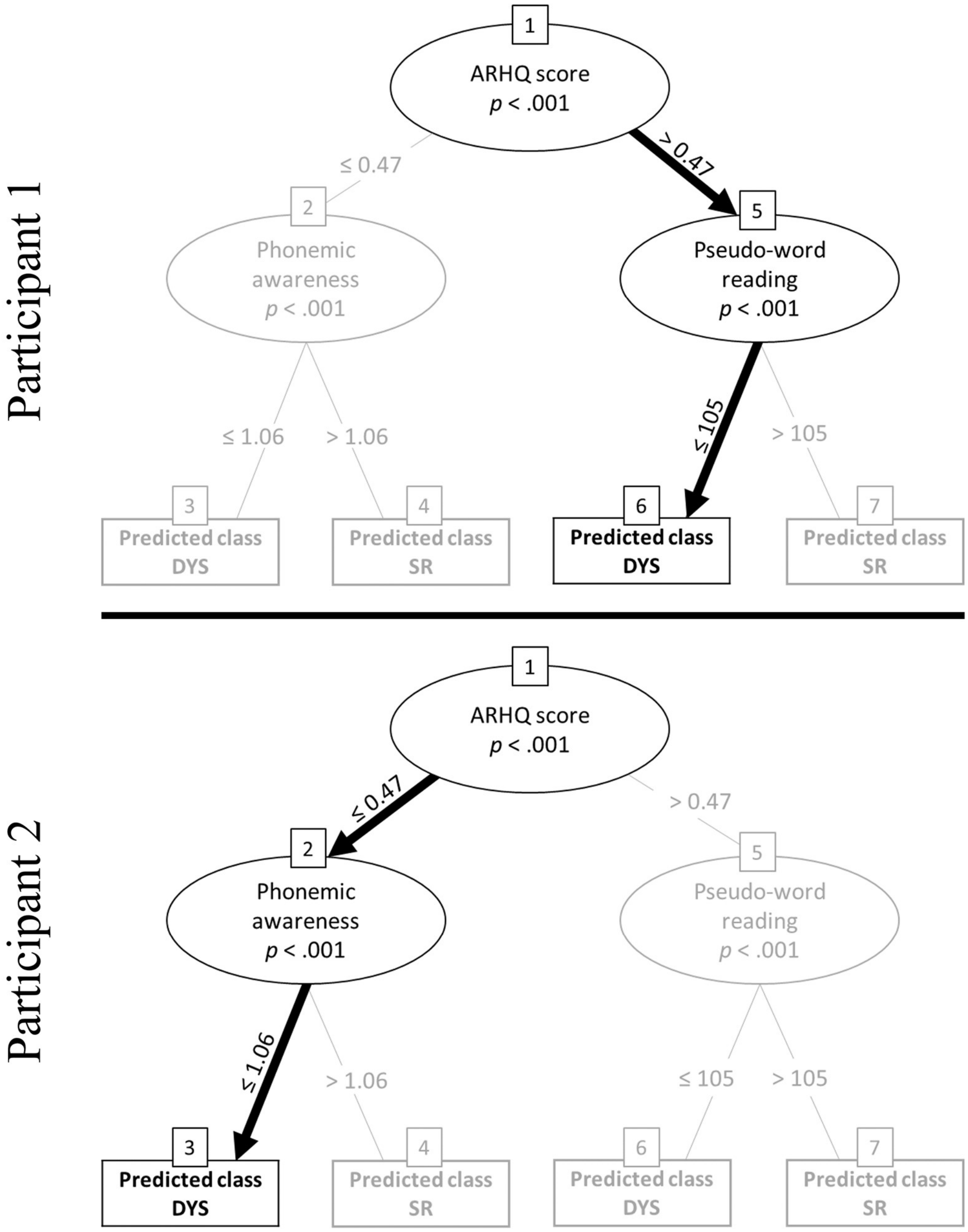

Figure 4.A. Illustration of how the classification procedures (with Tree 1) might be used for screening by clinicians. The Figure presents two examples (Participant 1, top part; Participant 2, bottom part). In each graph, bold arrows represent the decision path with respect to the obtained scores (see Table 3) and the conditional cut-off scores.

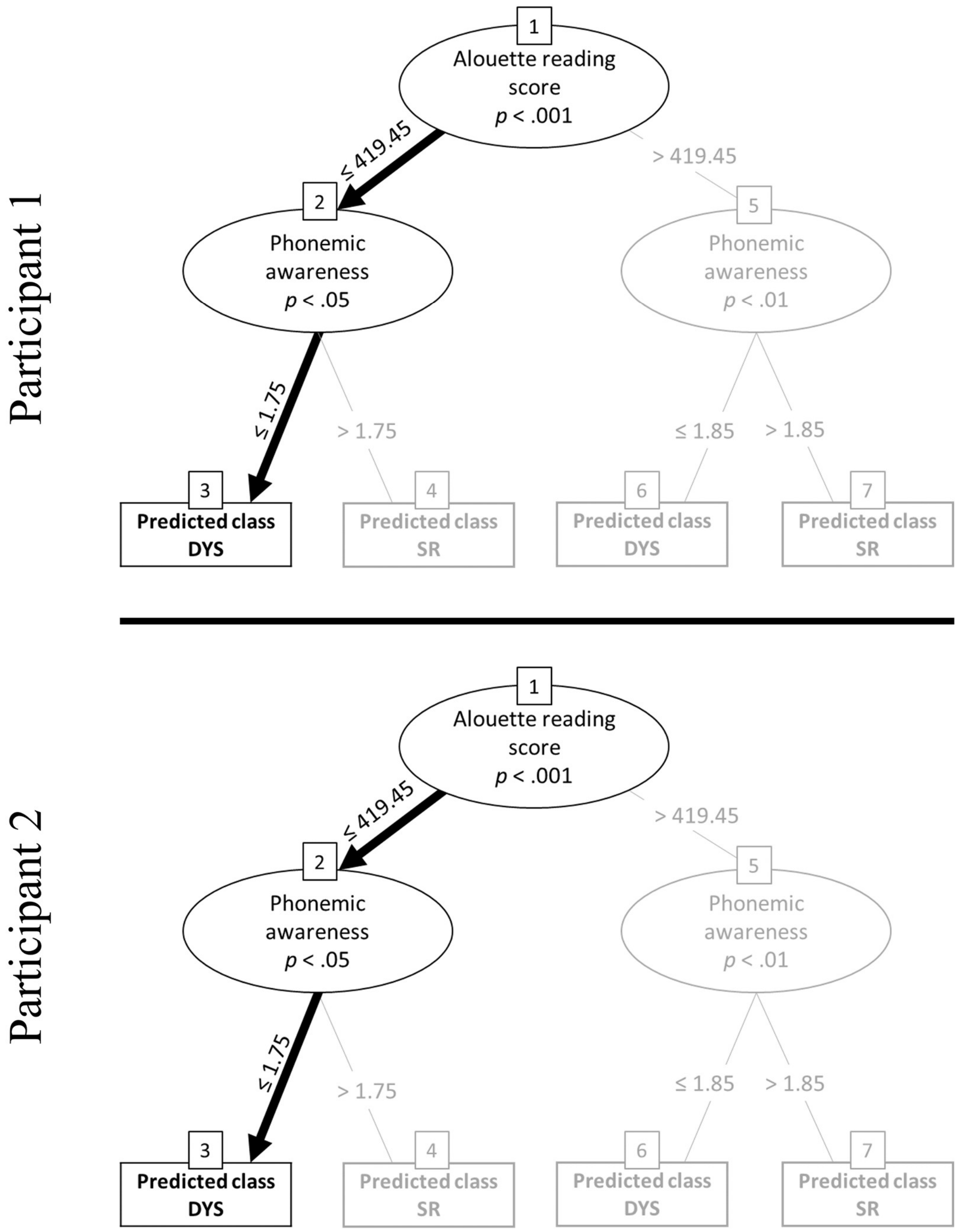

Figure 4.B. Illustration of how the classification procedures (with Tree 2) might be used for screening by clinicians. The Figure presents two examples (Participant 1, top part; Participant 2, bottom part). In each graph, bold arrows represent the decision path with respect to the obtained scores (see Table 3) and the conditional cut-off scores.